THE ANNALS
of
MATHEMATICAL
STATISTICS

Contents

ELIMINATION OF RANDOMIZATION IN CERTAIN STATISTICAL DECISION PROCEDURES AND ZERO-SUM TWO-PERSON GAMES[1]

By A. Dvoretzky,[2] A. Wald,[3] and J. Wolfowitz[3]

Institute for Numerical Analysis and Columbia University

Summary. The general existence of minimax strategies and other important properties proved in the theory of statistical decision functions (e.g., [3]) and the theory of games (e.g., [5]) depends upon the convexity of the space of decision functions and the convexity of the space of strategies. This convexity can be obtained by the use of randomized decision functions and mixed (randomized) strategies. In Section 2 of the present paper the authors state the extension (first announced in [1]) of a measure theoretical result known as Lyapunov's theorem [2]. This result is applied in Section 3 to the statistical decision problem where the number of distributions and decisions is finite. It is proved that when the distributions are continuous (more generally, "atomless," see footnote 7 below) randomization is unnecessary in the sense that every randomized decision function can be replaced by an equivalent nonrandomized decision function. Section 4 extends this result to the case when the decision space is compact. Section 5 extends the results of Section 3 to the sequential case. Sections 6 and 7 show, by counterexamples, that the results of Section 3 cannot be extended to the case of infinitely many distributions without new restrictions.[4] Section 8 gives sufficient conditions for the elimination of randomization under maintenance of ϵ-equivalence. Section 9 concludes with a restatement of the results in the language of the theory of games.

1. Introduction. We shall consider the following statistical decision problem: Let x be the generic point in an n-dimensional Euclidean[5] space R, and let Ω be a given class of cumulative distribution functions $F(x)$ in R. The cumulative distribution function $F(x)$ of the vector chance variable $X = (X_1, \cdots, X_n)$ with range in R is not known. It is known, however, that F is an element of the given class Ω. There is also given a space D whose elements d represent the possible decisions that can be made by the statistician in the problem under consideration. Let $W(F, d, x)$ denote the "loss" when F is the true distribution of

---

[1] The main results of this paper were announced without proof in an earlier publication [1] of the authors.

[2] On leave of absence from the Hebrew University, Jerusalem, Israel.

[3] Research under a contract with the Office of Naval Research.

[4] The impossibility of such an extension is related to the failure of Lyapunov's theorem when infinitely many measures are considered. (cf. A. Lyapunov, "Sur les fonctions-vecteurs complètement additives," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 10 (1946), pp. 277-279.)

[5] The restriction to a Euclidean space is not essential (see [1]).

X, the decision d is made and x is the observed value of X. We shall define the distance between two elements d_1 and d_2 of D by

$$(1.1) \qquad \rho(d_1, d_2) = \operatorname*{Sup}_{F,x} | W(F, d_1, x) - W(F, d_2, x) |.$$

Let B be the smallest Borel field of subsets of D which contains all open subsets of D as elements. Let B_0 be the totality of Borel sets of R. We shall assume that $W(F, d, x)$ is bounded[6] and, for every F, a function of d and x which is measurable $(B \times B_0)$. By a decision function $\delta(x)$ we mean a function which associates with each x a probability measure on D defined for all elements of B. We shall occasionally use the symbol δ_x instead of $\delta(x)$ when we want to emphasize that x is kept fixed. A decision function $\delta(x)$ is said to be nonrandomized if for every x the probability measure $\delta(x)$ assigns the probability one to a single point d of D. For any measurable subset D^* of D (D^* an element of B), the symbol $\delta(D^* \mid x)$ will denote the probability measure of D^* according to the set function $\delta(x)$. It will be assumed throughout this paper that for any given D^* the function $\delta(D^* \mid x)$ is a Borel measurable function of x. The adoption of a decision function $\delta(x)$ by the statistician means that he proceeds according to the following rule: Let x be the observed value of X. The element d of the space D is selected by an independent chance mechanism constructed in such a way that for any measurable subset D^* of D the probability that the selected element d will be included in D^* is equal to $\delta(D^* \mid x)$.

Given the sample point x and given that $\delta(x)$ is the decision function adopted, the expected value of the loss $W(F, d, x)$ is given by

$$(1.2) \qquad W^*(F, \delta, x) = \int_D W(F, d, x) \, d\delta_x.$$

The expected value of the loss $W(F, d, x)$ when F is the true distribution of X and $\delta(x)$ is the decision function adopted (but x is not known) is obviously equal to

$$(1.3) \qquad r(F, \delta) = \int_R W^*(F, \delta, x) \, dF(x).$$

The above expression is called the risk when F is true and δ is adopted.

We shall say that the decision functions $\delta(x)$ and $\delta^*(x)$ are equivalent if

$$(1.4) \qquad r(F, \delta^*) = r(F, \delta) \qquad \text{for all } F \text{ in } \Omega.$$

We shall say that $\delta(x)$ and $\delta^*(x)$ are strongly equivalent if for every measurable subset D^* of D we have

$$(1.5) \qquad \int_R \delta(D^* \mid x) \, dF(x) = \int_R \delta^*(D^* \mid x) \, dF(x) \qquad \text{for all } F \text{ in } \Omega.$$

---

[6] The restriction of boundedness is not essential (see [1]).

If δ and δ^* are strongly equivalent, they are equivalent for any loss function which is a function of F and d only.

For any positive ϵ, we shall say that $\delta(x)$ and $\delta^*(x)$ are ϵ-equivalent if

$$(1.6) \qquad | r(F, \delta) - r(F, \delta^*) | \leq \epsilon \qquad \text{for all } F \text{ in } \Omega,$$

and strongly ϵ-equivalent if

$$(1.7) \qquad \left| \int_R \delta(D^* \mid x) \, dF(x) - \int_R \delta^*(D^* \mid x) \, dF(x) \right| \leq \epsilon$$

for all measurable D^* and for all F in Ω.

In Section 2 we state a measure-theoretical result first announced in [1] and proved in [6]. This result is then used in Section 3 to prove that for every decision function there exists an equivalent, as well as a strongly equivalent, nonrandomized decision function δ^*, if Ω and D are finite and if each element $F(x)$ of Ω is atomless.[7] This result is extended in Section 4 to the case where D is compact. Section 5 deals with the sequential case for which similar results are proved. A precise definition of a sequential decision function is given in Section 5.

The finiteness of Ω is essential for the validity of the results given in Sections 2–5. The examples given in Section 6 show that even when Ω is such a simple class as the class of all univariate normal distributions with unit variance, there exist decision functions δ such that no equivalent nonrandomized decision functions exist. In Section 7, an example is given where a decision function δ and a positive ϵ exist such that no nonrandomized decision function δ^* is ϵ-equivalent to δ.

In Section 8, sufficient conditions are given which guarantee that for every δ and for every $\epsilon > 0$ there exists a nonrandomized decision function δ^* which is ϵ-equivalent to δ.

2. A measure-theoretical result. Let $\{y\} = Y$ be any space and let $\{S\} = \mathfrak{S}$ be a Borel field of subsets of Y. Let $\mu_k(S)(k = 1, \cdots, q)$ be a finite number of real-valued, σ-finite and countably additive set functions defined for all $S \epsilon \mathfrak{S}$. The following theorem was stated by the authors [1]:

THEOREM 2.1. *Let $\delta_j(y)$ $(j = 1, 2, \cdots, m)$ be real non-negative $\mathfrak{S}$-measurable functions satisfying*

$$(2.1) \qquad \sum_{j=1}^m \delta_j(y) = 1$$

for all $y \epsilon Y$. Then if the set functions $\mu_k(S)$ are atomless there exists a decomposition of Y into m disjoint subsets $S_1, \cdots, S_m$ belonging to $\mathfrak{S}$ having the property

[7] A set function μ defined on a Borel field $\mathfrak{S}$ is called atomless if it has the following property: If for some $S \epsilon \mathfrak{S}$, $\mu(S) \neq 0$, then there exists an $S' \subset S$ such that $S' \epsilon \mathfrak{S}$ and such that $\mu(S') \neq \mu(S)$ and $\mu(S') \neq 0$. A cumulative distribution function is called atomless if its associated set function is atomless.

that

$$(2.2) \qquad \int_Y \delta_j(y) \, d\mu_k(y) = \mu_k(S_j) \qquad\qquad (j = 1, \cdots, m; k = 1, \cdots, q).$$

If $\delta_j^*(y) = 1$ for all $y \, \epsilon \, S_j$ and $= 0$ for any other $y(j = 1, \cdots, m)$, then the above equation can be written as

$$(2.3) \qquad \int_Y \delta_j(y) \, d\mu_k(y) = \int_Y \delta_j^*(y) \, d\mu_k(y) \qquad (j = 1, \cdots, m; k = 1, \cdots, q).$$

This theorem is an extension of a result of A. Lyapunov [2] and is basic for deriving most of the results of the present paper.

3. Elimination of randomization when Ω and D are finite and each element $F(x)$ of Ω is atomless. In this section we shall assume that Ω consists of the elements $F_1(x), \cdots, F_p(x)$ and D of the elements $d_1, \cdots, d_m$. Moreover, we assume that $F_i(x)$ is atomless for $i = 1, \cdots, p$. A decision function $\delta(x)$ is now given by a vector function $\delta(x) = [\delta_1(x), \cdots, \delta_m(x)]$ such that

$$(3.1) \qquad\qquad \delta_j(x) \geqq 0, \qquad \sum_{j=1}^m \delta_j(x) = 1$$

for all $x \, \epsilon \, R$. Here $\delta_j(x)$ is the probability that the decision d_j will be made when x is the observed value of X. The risk when F_i is true and the decision function $\delta(x)$ is adopted is now given by

$$(3.2) \qquad\qquad r(F_i, \delta) = \sum_{j=1}^m \int_R W(F_i, d_j, x)\delta_j(x) \, dF_i(x).$$

A nonrandomized decision function $\delta^*(x)$ is a vector function whose components $\delta_j^*(x)$ can take only the values 0 and 1 for all x.

For any measurable subset S of R let

$$(3.3) \qquad \nu_{ij}(S) = \int_S W(F_i, d_j, x) \, dF_i(x) \qquad\qquad (i = 1, \cdots p; j = 1, \cdots, m).$$

Then the measures $\nu_{ij}(S)$ are finite, atomless and countably additive. Using these set functions, equation (3.2) can be written as

$$(3.4) \qquad\qquad r(F_i, \delta) = \sum_{j=1}^m \int_R \delta_j(x) \, d\nu_{ij}(x).$$

Replacing in Theorem 2.1 the space Y by R, the set of measures $\{\mu_1, \cdots, \mu_q\}$ by the set $\{\nu_{ij}\}(i = 1, \cdots, p; j = 1, \cdots, m)$, it follows from Theorem 2.1 that there exists a nonrandomized decision function $\delta^*(x)$ such that

$$(3.5) \qquad \int_R \delta_j(x) \, d\nu_{ij}(x) = \int_R \delta_j^*(x) \, d\nu_{ij}(x) \qquad (i = 1, \cdots, p; j = 1, \cdots, m).$$

This immediately yields the following theorems:

THEOREM 3.1. *If Ω and D are finite and if each element $F(x)$ of Ω is atomless, then for any decision function $\delta(x)$ there exists an equivalent nonrandomized decision function $\delta^*(x)$.*

Putting $W(F, d, x) = 1$ identically in F, d and x, equation (3.5) immediately yields the following theorem:

THEOREM 3.2. *If Ω and D are finite and if each element $F(x)$ of Ω is atomless, then for any decision function $\delta(x)$ there exists a strongly equivalent nonrandomized decision function $\delta^*(x)$.*

4. Elimination of randomization when Ω is finite, D is compact and each element $F(x)$ of Ω is atomless. Again, let $\Omega = \{F_1, \cdots, F_p\}$ where the distributions F are atomless. If the loss $W(F, d, x)$ does not depend on x, the finiteness of Ω implies that D is at least conditionally compact with respect to the metric (1.1) (see Theorem 3.1 in [3]). We postulate that D is compact (but permit the loss to depend on x), and shall prove that if $\delta(x)$ is any decision function, there exists a nonrandomized decision function $\delta^*(x)$ such that $\delta^*(x)$ is equivalent to $\delta(x)$, i.e.,

$$(4.1) \qquad\qquad r_i(\delta) = r_i(\delta^*) \qquad\qquad (i = 1, \cdots, p),$$

where $r_i(\delta)$ stands for $r(F_i, \delta)$.

Since D is compact there exists an infinite sequence of decompositions of the space D into a finite number of disjoint nonempty measurable sets, the lth decomposition to be $C(1, 1, \cdots, 1), \cdots, C(k_1, \cdots, k_l)$ with the properties:

(a) Any two sets C which have the same number of indices not all identical, are disjoint.

(b) The sum of all sets with the same number l of indices is D ($l = 1, 2, \cdots$ ad inf.).

(c) If the sequence of indices of one set C constitutes a proper initial part of the sequence of indices of another set C, the first set includes the second.

(d) The diameters of all sets with l indices are bounded above by $h(l)$ and

$$\lim_{l \to \infty} h(l) = 0.$$

Let l be fixed and define

$$(4.2) \qquad\qquad \Delta_{m_1, \cdots, m_l}(x) = \delta[C(m_1, \cdots, m_l \mid x].$$

Define, furthermore,

$$(4.3) \qquad W_i[x, C(m_1, \cdots, m_l)] = \frac{1}{\Delta_{m_1 \cdots m_l}(x)} \int_{C(m_1, \cdots, m_l)} W(F_i, d, x) \, d\delta_x$$

$$\text{if } \Delta_{m_1 \cdots m_l}(x) > 0,$$

$$= 0 \qquad \text{if } \Delta_{m_1 \cdots m_l}(x) = 0.$$

Clearly,

$$(4.4) \qquad r_i(\delta) = \sum_{m_l=1}^{k_l} \cdots \sum_{m_1=1}^{k_1} \int_R W_i[x, C(m_1, \cdots, m_l)]\Delta_{m_1\cdots m_l}(x) \, dF_i(x).$$

Considering a decision space D_i with elements $d_{m_1\cdots m_l}$ ($m_i = 1, \cdots, k_i$; $i = 1, \cdots, l$) and putting the loss $W(F_i, d_{m_1\cdots m_l}, x) = W_i[x, C(m_1, \cdots, m_l)]$, equations (3.3) and (3.5) imply that there exists a finite sequence of measurable functions $\bar{\Delta}_{m_1\cdots m_l}(x)$ ($m_1 = 1, \cdots, k_1$; $\cdots$; $m_l = 1, \cdots, k_l$) such that

$$(4.5) \qquad\qquad \bar{\Delta}_{m_1\cdots m_l}(x) = 0 \text{ or } 1 \qquad\qquad \text{for all } x,$$

$$(4.6) \qquad\qquad \sum_{m_l} \cdots \sum_{m_1} \bar{\Delta}_{m_1\cdots m_l}(x) = 1 \qquad\qquad \text{for all } x,$$

$$(4.7) \qquad\qquad \bar{\Delta}_{m_1\cdots m_l}(x) = 0 \qquad\qquad \text{whenever } \Delta_{m_1\cdots m_l}(x) = 0,$$

and

$$(4.8) \qquad \begin{aligned} \int_R & W_i[x, C(m_1, \cdots, m_l)]\bar{\Delta}_{m_1\cdots m_l}(x) \, dF_i(x) \\ &= \int_R W_i[x, C(m_1, \cdots, m_l)\Delta_{m_1\cdots m_l}(x) \, dF_i(x). \end{aligned}$$

Let now $\bar{\delta}(x)$ be the decision function for which

$$(4.9) \qquad\qquad \bar{\delta}[C(m_1, \cdots, m_l) \mid x] = \bar{\Delta}_{m_1\cdots m_l}(x)$$

and for any measurable subset $D_{m_1\cdots m_l}$ of $C(m_1, \cdots, m_l)$

$$(4.10) \qquad \bar{\delta}[D_{m_1\cdots m_l} \mid x]\bar{\Delta}_{m_1\cdots m_l}(x) = \frac{\delta(D_{m_1\cdots m_l} \mid x)}{\delta[C(m_1, \cdots, m_l) \mid x]},$$

where $\dfrac{\delta(D_{m_1\cdots m_l} \mid x)}{\delta[C(m_1,\cdots,m_l)\mid x]}$ is defined to be $= 0$ when $\delta[C(m_1, \cdots, m_l) \mid x] = 0$.

It then follows from (4.4) and (4.8) that

$$(4.11) \qquad\qquad r_i(\delta) = r_i(\bar{\delta}).$$

Applying the above result for $l = 1$, we conclude that there exists a decision function $\delta^1(x)$ with the following properties: The choice among the C's with one index is nonrandom. The decision, once given the C (with one index) chosen, is made according to $\delta(x)$. We have $\delta^1[C(m_1) \mid x] = 0$ whenever $\delta[C(m_1) \mid x] = 0$ and

$$r_i(\delta) = r_i(\delta^1) \qquad\qquad (i = 1, \cdots, p).$$

Repeat the above procedure for every C with two indices, using $W_i\{x, C(m_1, m_2)\}$ as weight function and $\delta^1(x)$ as the decision function. We

conclude that there exists a decision function $\delta^2(x)$ with the following properties: The choice among the C's with two indices is nonrandom. $\delta^2[C(m_1, m_2) \mid x] = 0$ whenever $\delta^1[C(m_1, m_2) \mid x] = 0$. The decision, once given the C (with two indices) chosen, is made according to $\delta^1(x)$ and, therefore, in accordance with $\delta(x)$. We have

$$\iint_R \int_{C(m_1)} W(F_i, d, x) \, d\delta_x^1 \, dF_i(x) = \iint_R \int_{C(m_1)} W(F_i, d, x) \, d\delta_x^2 \, dF_i(x) \quad \begin{matrix} (m_1 = 1, 2, \cdots, k_1) \\ (i = 1, \cdots, p). \end{matrix}$$

Repeat the above procedure for all C's with l indices, $l = 3, 4, \cdots$ ad inf. At the lth stage we obtain a decision function $\delta^l(x)$ with the following properties: The decision among the C's with l indices is nonrandom. $\delta^l[C(m_1, \cdots, m_l) \mid x] = 0$ whenever $\delta^{l-1}[C(m_1, \cdots, m_l) \mid x] = 0$. The decision, once given the chosen C with l indices, is made according to $\delta(x)$. We have

$$\iint_R \int_{C(m_1, \cdots, m_{l-1})} W(F_i, d, x) \, d\delta_x^{l-1} \, dF_i(x) = \iint_R \int_{C(m_1, \cdots, m_{l-1})} W(F_i, d, x) \, d\delta_x^l \, dF_i(x)$$

$$\begin{pmatrix} i = 1, \cdots, p \\ m_1 = 1, \cdots, k_1 \\ m_{l-1} = 1, \cdots, k_{l-1} \end{pmatrix}.$$

Hold x fixed and let $C(x; l)$ be that C with l indices for which

$$\int_{C(x;l)} d\delta_x^l = 1.$$

Then $C(x; l + 1)$ is a proper subset of $C(x; l)$ for every positive l. The sequence $C(x; l)$, $l = 1, 2, \cdots$, determines, because D is compact, a unique limit point $c(x)$ such that any neighborhood of $c(x)$ contains almost all sets $C(x; l)$. Hence the sequence of probability measures $\delta_x^l (l = 1, 2, \cdots,$ ad inf.) converges to a limit probability measure δ_x^* which assigns probability one to any measurable set which contains the point $c(x)$. Since $W(F_i, d, x)$ is continuous in d, we have

$$(4.12) \qquad \lim_{l \to \infty} \int_D W(F_i, d, x) \, d\delta_x^l = \int_D W(F_i, d, x) \, d\delta_x^*$$

for any x.

Now let x vary over R. It follows from (4.12) and the boundedness of $W(F, d, x)$ that $\lim_{l \to \infty} r_i(\delta^l) = r_i(\delta^*)$. Since $r_i(\delta^l) = r_i(\delta)$, also $r_i(\delta^*) = r_i(\delta)$ $(i = 1, \cdots, p)$. Thus the probability measures $\delta^*(x)$ constitute the desired nonrandomized decision function.

It remains to show that for any measurable subset D^* of D, the function $\delta^*(D^* \mid x)$ is a measurable function of x. The measurability of $\delta^*(D^* \mid x)$ can easily be shown for any D^*, if it is shown for all closed sets D^*, since every measurable set can be attained by a denumerable number of Borel operations (denumerably infinite sums and complements) starting with closed sets. Thus

we shall assume that D^* is closed. For any positive ρ let D_ρ^* be the sum of all open spheres with center in D^* and radius ρ. It is easy to see that

$$\delta^*(D_{2\rho}^* \mid x) \geqq \liminf_{l \to \infty} \delta^l(D_\rho^* \mid x) \geqq \delta^*(D^* \mid x).$$

Since $\lim_{\rho \to 0} \delta^*(D_{2\rho}^* \mid x) = \delta^*(D^* \mid x)$, it follows from the above relation that

$$\lim_{\rho \to 0} \liminf_l \delta^l(D_\rho^* \mid x) = \delta^*(D^* \mid x).$$

Since $\delta^l(D_\rho^* \mid x)$ is a measurable function of x, the measurability of $\delta^*(D^* \mid x)$ is proved.

5. Elimination of randomization in the sequential case. In this section we shall consider the following sequential decision problem: Let $X = \{X_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ be a sequence of chance variables. Let x be the generic point in the space R of all infinite sequences of real numbers, i.e., $x = \{x_n\}$ $(n = 1, 2, \cdots, \text{ad inf.})$ where each x_n is a real number. It is known that the distribution function $F(x)$ of X is an element of Ω, where Ω consists of a finite number of distribution functions $F_1(x), \cdots, F_p(x)$, and that the distribution function of X_1 is continuous according to $F_i(x)$, $i = 1, \cdots, p$. The statistician is assumed to have a choice of a finite number of (terminal) decisions $d_1, \cdots, d_m$, i.e., the space D consists of the elements $d_1, d_2, \cdots, d_m$. A decision rule δ is now given by a sequence of nonnegative, measurable functions $\delta_{\nu t}(x_1, \cdots, x_t)$ $(\nu = 0, 1, \cdots, m; t = 1, 2, \cdots, \text{ad inf.})$ satisfying

(5.1) $$\sum_{\nu=0}^m \delta_{\nu t}(x_1, \cdots, x_t) = 1$$

for $-\infty < x_1, \cdots, x_t < \infty$. The decision rule δ is defined in terms of the functions $\delta_{\nu t}$ as follows: After the value x_1 of X_1 has been observed, the statistician decides either to continue experimentation and take another observation, or to stop further experimentation and adopt a terminal decision $d_j(j = 1, \cdots, m)$ with the respective probabilities $\delta_{01}(x_1)$ and $\delta_{j1}(x_1)$ $(j = 1, \cdots, m)$. If it is decided to continue experimentation, a value x_2 of X_2 is observed and it is again decided either to take a further observation or adopt a terminal decision $d_j(j = 1, \cdots, m)$ with the respective probabilities $\delta_{02}(x_1, x_2)$ and $\delta_{j2}(x_1, x_2)(j = 1, \cdots, m)$, etc. The decision rule is called nonrandomized if each $\delta_{\nu t}$ can take only the values 0 and 1.

Let $v_{i\nu t}(x_1, \cdots, x_t)$ represent the sum of the loss and the cost of experimentation when F_i is true, the terminal decision d_ν is made and experimentation is terminated with the tth observation

$$(\nu = 1, 2, \cdots, m; i = 1, \cdots, p; t = 1, 2, \cdots, \text{ad inf.}).$$

The functions $v_{i\nu t}(x_1, \cdots, x_t)$ are assumed to be finite, nonnegative and measurable. We shall consider only decision rules δ for which the probability is one that experimentation will be terminated at some finite stage. The risk (ex-

pected loss plus expected cost of experimentation) when F_i is true and the rule δ is adopted is then given by

$$(5.2) \quad r_i(\delta) = \sum_{t=1}^{\infty} \sum_{\nu=1}^{m} \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t) \delta_{01}(x_1) \delta_{02}(x_1, x_2) \cdots \delta_{0(t-1)}(x_1, \cdots, x_{t-1})$$
$$\cdot \delta_{\nu t}(x_1, \cdots, x_t) \, dF_{it}(x_1, \cdots, x_t),$$

where R_t is the t-dimensional space of $x_1, \cdots, x_t$ and $F_{it}(x_1, \cdots, x_t)$ is the cumulative distribution function of $X_1, \cdots, X_t$ when F_i is the distribution function of X.

We shall say that the decision rules δ^1 and δ^2 are equivalent if $r_i(\delta^1) = r_i(\delta^2)$ for $i = 1, \cdots, p$. We shall say that δ^1 and δ^2 are strongly equivalent if

$$(5.3) \quad \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t) \delta^1_{01}(x_1) \cdots \delta^1_{0(t-1)}(x_1, \cdots, x_{t-1}) \delta^1_{\nu t}(x_1, \cdots, x_t) \, dF_{it}$$
$$= \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t) \delta^2_{01}(x_1) \cdots \delta^2_{0(t-1)}(x_1, \cdots, x_{t-1}) \delta^2_{\nu t}(x_1, \cdots, x_t) \, dF_{it}$$

for $i = 1, 2, \cdots, p$; $\nu = 1, \cdots, m$ and $t = 1, 2, \cdots,$ ad inf.

Clearly, if δ^1 and δ^2 are strongly equivalent and if the functions $v_{i\nu t}(x_1, \cdots, x_t)$ reduce to constants $v_{i\nu t}$, then δ^1 and δ^2 are equivalent for all possible choices of the constants $v_{i\nu t}$.

Let

$$(5.4) \quad \varphi_i(x, \delta) =$$
$$\sum_{t=1}^{\infty} \sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t) \delta_{01}(x_1) \cdots \delta_{0(t-1)}(x_1, \cdots, x_{t-1}) \delta_{\nu t}(x_1, \cdots, x_t).$$

We shall prove the following lemma:

LEMMA 5.1. *Let δ be a decision rule for which $\varphi_i(x, \delta) < \infty$ for all x, except perhaps on a set of x's whose probability is zero according to every distribution function $F_i(x)(i = 1, \cdots, p)$. Let τ and T be given positive integers. Then there exists a decision function $\bar{\delta}$ with the following properties:*

$$(5.5) \quad \bar{\delta}_{\nu\tau}(x_1, \cdots, x_\tau) = 0 \quad \text{or} \quad 1, \quad \sum_{\nu=0}^{m} \bar{\delta}_{\nu\tau}(x_1, \cdots, x_\tau) = 1,$$

for every point in $R_\tau(\nu = 0, 1, \cdots, m)$,

$$(5.6) \quad \bar{\delta}_{\nu t}(x_1, \cdots, x_t) = \delta_{\nu t}(x_1, \cdots, x_t) \qquad (\nu = 0, 1, \cdots, m; t \neq \tau),$$

$$(5.7) \quad r_i(\delta) = r_i(\bar{\delta}) \qquad\qquad (i = 1, \cdots, p),$$

$$(5.8) \quad \int_{R_t} v_{i\nu t} \delta_{01} \cdots \delta_{0(t-1)} \delta_{\nu t} \, dF_{it} = \int_{R_t} v_{i\nu t} \bar{\delta}_{01} \cdots \bar{\delta}_{0(t-1)} \bar{\delta}_{\nu t} \, dF_{it}$$
$$(\nu = 1, \cdots, m; t = 1, \cdots, T),$$

$$(5.9) \quad \varphi_i(x, \bar{\delta}) < \infty,$$

for all x except perhaps on a set whose probability is zero according to every distribution $F_i(x)(i = 1, \cdots, p)$.

PROOF. We can write $\varphi_i(x, \delta)$ as follows:

$$(5.10) \qquad \varphi_i(x, \delta) = \sum_{t=1}^{\tau-1} \sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t}$$

$$+ \sum_{t=\tau}^{\infty} \sum_{\nu=0}^{m} g_{i\nu\tau t}(x_1, \cdots, x_t)\delta_{\tau\tau},$$

where $g_{i\nu\tau t}(x_1, \cdots, x_t)$ does not depend on $\delta_{0\tau}, \delta_{1\tau}, \cdots, \delta_{m\tau}$. The first double sum reduces to zero when $\tau = 1$. Clearly, if a $\bar\delta$ with the desired properties exists, then

$$(5.11) \qquad \varphi_i(x, \bar\delta) = \sum_{t=1}^{\tau-1} \sum_{\nu=1}^{m} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t}$$

$$+ \sum_{t=\tau}^{\infty} \sum_{\nu=0}^{m} g_{i\nu\tau t}(x_1, \cdots, x_t)\bar\delta_{\tau\tau}.$$

For any subset S of R, let

$$(5.12) \qquad \mu_{i\nu\tau t}(S) = \int_S g_{i\nu\tau t}(x_1, \cdots, x_t)\, dF_i \qquad (t = \tau, \tau+1, \cdots, T),$$

and

$$(5.13) \qquad \mu_{i\nu\tau}(S) = \int_S \left[\sum_{t=T+1}^{\infty} g_{i\nu\tau t}(x_1, \cdots x_t) \right] dF_i.$$

The measures $\mu_{i\nu\tau t}$ are not defined if $\tau > T$. Clearly, the measures

$$\mu_{i\nu\tau t}(\nu = 0, 1, \cdots, m; t = \tau, \tau+1, \cdots, T)$$

and the measures $\mu_{i\nu\tau}(\nu = 1, \cdots, m)$ are nonnegative, countably additive and σ-finite. Since for any x for which $\varphi_i(x, \delta) < \infty$ and $\delta_{0\tau} > 0$, the sum

$$\sum_{t=T+1}^{\infty} g_{i0\tau t}(x_1, \cdots, x_t) < \infty,$$

it follows from the assumptions of Lemma 5.1 that $\mu_{i0\tau}$ is σ-finite over the space R' consisting of all x for which $\delta_{0\tau} > 0$. Of course, $\mu_{i0\tau}$ is nonnegative and countably additive. Let R'' be the set of all points x for which $\delta_{0\tau} = 0$. We put

$$(5.14) \qquad \bar\delta_{0\tau}(x_1, \cdots, x_\tau) = 0 \quad \text{for all} \quad x \text{ in } R''.$$

Application of Theorem 2.1 to each of the spaces R' and R'' shows that there exist measurable functions $\bar\delta_{\nu\tau}(x_1, \cdots, x_\tau)(\nu = 0, 1, \cdots, m)$ such that in addition to (5.14) the following conditions hold:

$$(5.15) \quad \bar\delta_{\nu\tau} = 0 \quad \text{or} \quad 1(\nu = 0, 1, \cdots m) \quad \text{and} \quad \sum_{\nu=0}^{m} \bar\delta_{\nu\tau} = 1 \quad \text{for all } x,$$

(5.16) $\qquad \int_R \delta_{rr}\, d\mu_{i\nu\tau i} = \int_R \bar{\delta}_{rr}\, d\mu_{i\nu\tau i}$

$$(i = 1, \cdots, p; \nu = 0, 1, \cdots m; t = \tau, \tau + 1, \cdots, T),$$

(5.17) $\qquad \int_R \delta_{rr}\, d\mu_{i\nu\tau} = \int_R \bar{\delta}_{rr}\, d\mu_{i\nu\tau} \qquad (i = 1, \cdots, p; \nu = 0, 1, \cdots m).$

Lemma 5.1 is a simple consequence of the equations (5.14)–(5.17).

For any positive integer u, we shall say that a decision rule δ is truncated at the uth stage if $\delta_{0u'} = 0$ for $u' \geqq u$ identically in x.

THEOREM 5.1. *If δ is truncated at the uth stage there exists a nonrandomized decision rule δ^* that is strongly equivalent to δ.*

PROOF. It is sufficient to prove Theorem 5.1 in the case where $\delta_{\nu t} = 0$ for $t > u$ and $\nu \neq 1$ and $\delta_{1t} = 1$ for $t > u$. Clearly, $\varphi_i(x, \delta) < \infty$ for all x. Putting $\tau = 1$ and $T = u$ in Lemma 5.1, this lemma implies the existence of a decision rule δ^1 with the following properties: (a) δ^1 is strongly equivalent to δ; (b) $\delta^1_{\nu 1} = 0$ or 1 ($\nu = 0, 1, \cdots, m$); (c) $\delta^1_{\nu t} = \delta_{\nu t}$ for $\nu = 0, 1, \cdots, m$ and $t > 1$. Applying Lemma 5.1 to δ^1 and putting $\tau = 2$ and $T = u$, we see that there exists a decision rule δ^2 with the following properties: (a) δ^2 is strongly equivalent to δ^1; (b) $\delta^2_{\nu 2} = 0$ or 1 ($\nu = 0, 1, \cdots, m$); (c) $\delta^2_{\nu t} = \delta^1_{\nu t}$ for $\nu = 0, 1, \cdots, m$ and $t \neq 2$. Continuing this procedure, at the uth step we obtain a decision rule δ^u that is nonrandomized and is strongly equivalent to all the preceding ones. This proves our theorem.

We shall say that two decision rules δ^1 and δ^2 are strongly equivalent up to the Tth stage if

(5.18)
$$\int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^1_{01} \cdots \delta^1_{0(t-1)}\delta^1_{\nu t}\, dF_{it}$$
$$= \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta^2_{01} \cdots \delta^2_{0(t-1)}\delta^2_{\nu t}\, dF_{it}$$

$$\text{for} \quad i = 1, \cdots, p; \nu = 1, \cdots, m \quad \text{and} \quad t = 1, \cdots, T.$$

Furthermore, we shall say that a decision rule δ is nonrandomized up to the stage T if $\delta_{\nu t} = 0$ or 1 for $\nu = 0, 1, \cdots, m$ and $t = 1, \cdots, T$.

We now prove the following theorem.

THEOREM 5.2. *If δ is a decision rule for which $\varphi_i(x, \delta) < \infty$, except perhaps on a set of x's of probability zero according to every $F_i(x) (i = 1, \cdots, p)$, then there exists a nonrandomized decision rule δ^* that is equivalent to δ.*

PROOF. Let $\{\epsilon_i\}$ and $\{\eta_i\} (i = 1, 2, \cdots, \text{ad inf.})$ be two sequences of positive numbers such that $\lim_{i=\infty} \epsilon_i = 0$ and $\lim_{i=\infty} \eta_i = \infty$. Let T_1 be a positive integer such that

(5.19) $\quad r_i(\delta) - \sum_{t=1}^{T_1} \sum_{\nu=1}^{m} \int_{R_t} v_{i\nu t}(x_1, \cdots, x_t)\delta_{01} \cdots \delta_{0(t-1)}\delta_{\nu t}\, dF_{it} < \epsilon_1 \quad \text{if} \quad r_i(\delta) < \infty,$

and

$$(5.20) \quad \sum_{i=1}^{T_1} \sum_{\nu=1}^{m} \int_{R_t} v_{\nu t}(x_1, \cdots, x_t) \delta_{01} \cdots \delta_{0(t-1)} \delta_{\nu t} \, dF_{it} > \eta_1 \quad \text{if} \quad r_i(\delta) = \infty.$$

Let δ^1 be a decision rule such that $\varphi_i(x, \delta^1) < \infty$ (except perhaps on a set of probability measure zero); δ^1 is equivalent to δ; δ^1 is strongly equivalent to δ up to the T_1^{th} stage; δ^1 is nonrandomized up to the T_1^{th} stage and $\delta^1_{\nu t} = \delta_{\nu t}$ for $t > T_1$. The existence of such a decision rule follows from a repeated application of Lemma 5.1. In general, after $\delta^1, \cdots, \delta^j$ and $T_1, \cdots, T_j$ are given, let δ^{j+1} be a decision rule such that $\varphi_i(x, \delta^{j+1}) < \infty$ (except perhaps on a set of probability measure zero); δ^{j+1} is equivalent to δ^j; δ^{j+1} is strongly equivalent to δ^j up to the T_{j+1}^{th} stage, where T_{j+1} is a positive integer chosen so that $T_{j+1} > T_j$ and (5.19) and (5.20) hold with δ replaced by δ^j, ϵ_1 replaced by ϵ_{j+1} and η_1 replaced by η_{j+1}; δ^{j+1} is nonrandomized up to the stage T_{j+1}; $\delta^{j+1}_{\nu t} = \delta^j_{\nu t}$ for $t \leq T_j$ and $\delta^{j+1}_{\nu t} = \delta_{\nu t}$ for $t > T_{j+1}$. The existence of such a decision rule δ^{j+1} follows again from a repeated application of Lemma 5.1.

Let δ^* be the decision rule given by the equations

$$(5.21) \qquad \delta^*_{\nu t} = \delta^t_{\nu t} \qquad (\nu = 0, 1, \cdots, m; t = 1, 2, \cdots, \quad \text{ad inf.}).$$

It follows easily from the above stated properties of the decision rules δ^j $(j = 1, 2, \cdots, \text{ad inf.})$ that δ^* is nonrandomized and $r_i(\delta^*) = r_i(\delta)(i = 1, \cdots, p)$. This completes the proof of Theorem 5.2.

6. Examples where admissible[3] decision functions do not admit equivalent nonrandomized decision functions.

In this section we shall construct examples which show that there exist admissible decision functions $\delta(x)$ which do not admit equivalent nonrandomized decision functions $\delta^*(x)$.

EXAMPLE 1. Let X be a normally distributed chance variable with unknown mean θ and variance unity. This means that Ω is the totality of all univariate normal distributions with unit variance. Suppose we wish to test the hypothesis H_0 that the true mean θ is rational on the basis of a single observation x on X. Thus, D consists of two elements d_1 and d_2 where d_1 is the decision to accept H_0 and d_2 is the decision to reject H_0. For any decision function $\delta(x)$, let $\delta_1(x)$ denote the value of $\delta(d_1 \mid x)$. Let the loss be zero when a correct decision is made, and the loss be one when a wrong decision is made. Then the risk when θ is the true mean and the decision function $\delta(x)$ is adopted is given by

$$(6.1) \qquad r(\theta, \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} \delta_1(x) \, dx \qquad \text{when } \theta \text{ is irrational,}$$

$$(6.2) \qquad r(\theta, \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} (1 - \delta_1(x)) \, dx \qquad \text{when } \theta \text{ is rational.}$$

---

[3] A decision function with risk function $r(F)$ is called admissible if there exists no other decision function with risk function $r'(F)$ such that $r'(F) \leq r(F)$ for every $F \in \Omega$, and the inequality sign holds for at least one $F \in \Omega$.

Let $\delta_1^0(x) = \frac{1}{2}$ for all x. Clearly,

$$(6.3) \qquad\qquad r(\theta, \delta^0) = \frac{1}{2}$$

for all θ. We shall now show that $\delta^0(x)$ is an admissible decision function. For suppose that there exists a decision function $\delta'(x)$ such that

$$(6.4) \qquad\qquad r(\theta, \delta') \leqq r(\theta, \delta^0) = \frac{1}{2}$$

for all θ, and

$$(6.5) \qquad\qquad r(\theta_1, \delta') < r(\theta_1, \delta^0) = \frac{1}{2}$$

for some value θ_1. Suppose first that θ_1 is rational. Since the integrals in (6.1) and (6.2) are continuous functions of θ, for an irrational value θ_2 sufficiently near to θ_1 we shall have $r(\theta_2, \delta') > \frac{1}{2}$ which contradicts (6.4). Thus, θ_1 cannot be rational. In a similar way, one can show that θ_1 cannot be irrational. Hence, the assumption that a decision function $\delta'(x)$ satisfying (6.4) and (6.5) exists leads to a contradiction and the admissibility of $\delta^0(x)$ is proved.

Let now $\delta^*(x)$ be any decision function for which

$$(6.6) \qquad\qquad r(\theta, \delta^*) = r(\theta, \delta^0)$$

for all θ. Now (6.6) implies that

$$(6.7) \qquad\qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} (\delta_1(x) - \delta_1^*(x))\, dx = 0$$

identically in θ. Since $\delta_1(x) - \delta_1^*(x)$ is a bounded function of x, it follows from the uniqueness properties of the Laplace transform that (6.7) can hold only if $\delta_1(x) - \delta_1^*(x) = 0$ except perhaps on a set of measure zero. Hence, no nonrandomized decision function $\delta^*(x)$ can satisfy (6.6).

In the above example, the distributions consistent with the hypothesis H_0 which is to be tested (normal distributions with rational means) are not well separated from the alternative distributions (normal distributions with irrational means). One might think that this is perhaps the reason for the existence of an admissible decision function δ^0 such that no nonrandomized decision function δ^* can have as good a risk function as δ^0 has. That this need not be so, is shown by the following:

EXAMPLE 2. Suppose that X is a normally distributed chance variable with mean θ and variance unity. The value of θ is unknown. It is known, however, that the true value of θ is contained in the union of the two intervals $[-2, -1]$ and $[1, 2]$. Suppose that we want to test the hypothesis that θ is contained in the interval $[-2, -1]$ on the basis of a single observation x on X. Suppose, furthermore, that the chance variable X itself is not observable and only the chance variable $Y = f(X)$ can be observed where $f(x) = x$ when $|x| < 1$, and $= |x|$ when $|x| \geqq 1$. Let the loss be zero when a correct decision is made, and one when a wrong decision is made. For any decision function $\delta(y)$, let

$\delta_1(y)$ denote the value of $\delta(d_1 \mid y)$ where d_1 denotes the decision to accept H_0. Let $\delta^0(y)$ be the following decision function:

$$
\begin{aligned}
\delta_1^0(y) &= 1 \quad \text{when} \quad -1 < y < 0 \\
&= 0 \quad \text{when} \quad 0 \leqq y < 1 \\
&= \tfrac{1}{2} \quad \text{when} \quad y \geqq 1.
\end{aligned}
$$

(6.8)

First we shall show that $\delta^0(y)$ is an admissible decision function. For this purpose, consider the following probability density function $g(\theta)$ in the parameter space: $g(\theta) = \tfrac{1}{2}$ when $-2 \leqq \theta \leqq -1$ or $1 \leqq \theta \leqq 2$, $= 0$ for all other θ. If we interpret $g(\theta)$ as the a priori probability distribution of θ, the a posteriori probability of the θ-interval $[-2, -1]$ is greater (less) than the a posteriori probability of the θ-interval $[1, 2]$ when $-1 < y < 0$ ($0 < y < 1$), and the a posteriori probabilities of the two intervals are equal to each other when $y = 0$ or $y \geqq 1$. Hence, $\delta^0(y)$ is a Bayes solution relative to the a priori distribution $g(\theta)$, i.e.,

$$
(6.9) \qquad \int_{2}^{-1} r(\theta, \delta^0) \, d\theta + \int_{1}^{2} r(\theta, \delta^0) \, d\theta \leqq \int_{2}^{-1} r(\theta, \delta) \, d\theta + \int_{1}^{2} r(\theta, \delta) \, d\theta
$$

for any decision function δ. Suppose now that δ is a decision function for which $r(\theta, \delta) \leqq r(\theta, \delta^0)$ for all θ. It then follows from (6.9) that $r(\theta, \delta) < r(\theta, \delta^0)$ can hold at most on a set of θ's of measure zero. Since, as can easily be verified, $r(\theta, \delta)$ and $r(\theta, \delta^0)$ are continuous functions of θ, it follows that $r(\theta, \delta) = r(\theta, \delta^0)$ everywhere and the admissibility of δ^0 is proved.

Let now $\delta'(y)$ be any decision function for which $r(\theta, \delta') = r(\theta, \delta^0)$ for all θ, i.e.,

$$
(6.10) \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} [\delta^0(y) - \delta'(y)] \, dx = 0 \qquad \text{for all } \theta.
$$

Since $\delta_1^0(y) - \delta_1'(y)$ is a bounded function of x, it follows from the uniqueness properties of the Laplace transform that (6.10) can hold only if $\delta_1^0(y) = \delta_1'(y)$ except perhaps on a set of measure zero. Thus, no nonrandomized decision function δ^* exists such that $r(\theta, \delta^*) = r(\theta, \delta^0)$ for all θ.

7. Compactness of Ω in the ordinary sense is not sufficient for the existence of ϵ-equivalent nonrandomized decision functions. Let $\Omega = \{F\}$ be the totality of density functions[9] on the interval $0 \leqq x \leqq 1$ for which $F(x) \leqq c$ for every x, where c is some positive constant greater than 2. The sample space will be the interval $0 \leqq x \leqq 1$. We shall say that the sequence $F_1, F_2, \cdots$ converges to F if

$$
\lim_{n \to \infty} \int_{-\infty}^{x} F_n(y) \, dy = \int_{-\infty}^{x} F(y) \, dy
$$

---

[9] Here $F(x)$ denotes a density function. This represents a change in notation from preceding sections.

for every real x. The set Ω is compact in the sense of the above convergence definition.[10] Let A be a fixed interval $a_1 \leqq x \leqq a_2$ where $0 < a_1 < a_2 < 1$. Let $D = \{d_1, d_2\}$ and define W as follows:

$$W(F, d_1) + W(F, d_2) \equiv 1,$$

$$W(F, d_1) = 0 \text{ or } 1$$

according as the probability of A under F is rational or not. For any decision function $\delta(x)$, let $\delta_1(x)$ denote the probability assigned to d_1 by $\delta(x)$, i.e., $\delta_1(x) = \delta(d_1 \mid x)$.

Let $\delta'(x)$ be the decision function for which $\delta_1'(x) \equiv \frac{1}{2}$. We shall prove that $\delta'(x)$ is an admissible decision function. For suppose there exists a decision function $\delta^0(x)$ such that

(7.1) $$r(F, \delta^0) \leqq r(F, \delta') = \frac{1}{2}$$

for every F, and for F_0 we have

(7.2) $$r(F_0, \delta^0) < r(F_0, \delta').$$

Now, if $F_i \to F_0$ and $W(F_i, d_1) = W(F_0, d_1)$ for every i, then $r(F_i, \delta) \to r(F_0, \delta)$ for every decision function $\delta(x)$, and, in particular, for $\delta^0(x)$. If $F_i \to F_0$ and $W(F_i, d_1) + W(F_0, d_1) = 1$ for every i, then $r(F_i, \delta) \to 1 - r(F_0, \delta)$ for every decision function $\delta(x)$ and, in particular, for $\delta^0(x)$. Clearly, we can construct two sequences of functions F such that each sequence converges to F_0, the probability of A according to every member of the first sequence is rational, and the probability of A according to every member of the second sequence is irrational. Because of (7.2) it follows that inequality (7.1) will be violated for almost every member of one of these two sequences. Hence δ' is admissible.

Let us now prove that there cannot exist a nonrandomized decision function $\delta^*(x)$ such that

(7.3) $$r(F, \delta^*) \leqq r(F, \delta') + \frac{1}{4} = \frac{3}{4}$$

for every $F \in \Omega$. Suppose there were such a decision function $\delta^*(x)$. Let H be the set of x's where $\delta_1^*(x) = 1$, and let $\bar{H}$ be the complement of H with respect to the interval $[0, 1]$. If H is a set of measure zero or one then obviously (7.3) is violated for some F. Thus, it is sufficient to consider the case when H is a set of positive measure $\alpha < 1$. Suppose for a moment that $\alpha > \frac{1}{2}$. Let G be the density which is zero on $\bar{H}$ and constant on H. From (7.3) it follows that $P\{A \mid G\}$ is rational. There exists a density $G' \in \Omega$ such that $P\{H \mid G'\} > \frac{3}{4}$ and $P\{A \mid G'\}$ is irrational. But then (7.3) is violated for G'. If $\alpha \leqq \frac{1}{2}$, let $\bar{G}$ be the density which is zero on H and constant on $\bar{H}$. From (7.3) it follows that $P\{A \mid \bar{G}\}$ is irrational. There exists a density $\bar{G}' \in \Omega$ such that $P\{\bar{H} \mid \bar{G}'\} >$

---

[10] The cumulative distribution functions are well-known to be compact in the usual convergence sense. Since the densities are bounded above the limit cumulative distribution function must be absolutely continuous.

$\frac{3}{4}$ and $P\{A \mid \bar{G}'\}$ is rational. But then (7.3) is violated for $\bar{G}'$. Thus (7.3) **can** never hold for every $F \, \epsilon \, \Omega$ and the desired result is proved.

8. Sufficient conditions for the existence of ϵ-equivalent nonrandomized decision functions. In this section we shall consider the nonsequential decision problem (as described in the introduction), and we shall give sufficient conditions for the existence of ϵ-equivalent nonrandomized decision functions. We shall consider the following four metrics in the space Ω:

$$(8.1) \qquad \rho_1(F_1, F_2) = \operatorname*{Sup}_{S} \left| \int_S dF_1 - \int_S dF_2 \right|$$

when S is any measurable subset of R,

$$(8.2) \qquad \rho_2(F_1, F_2) = \operatorname*{Sup}_{d,x} \left| W(F_1, d, x) - W(F_2, d, x) \right|,$$

$$(8.3) \qquad \rho_3(F_1, F_2) = \rho_1(F_1, F_2) + \rho_2(F_1, F_2),$$

$$(8.4) \qquad \rho_4(F_1, F_2) = \operatorname*{Sup}_{\delta} \left| r(F_1, \delta) - r(F_2, \delta) \right|.$$

First we prove the following lemma:

LEMMA 8.1. *If Ω is conditionally compact in the sense of the metric ρ_3, then it is conditionally compact in the sense of the metric ρ_4.*

PROOF. Let $\{F_i\}(i = 1, 2, \cdots$, ad inf.) be a Cauchy sequence in the sense of the metric ρ_3, i.e.,

$$(8.5) \qquad \lim_{i,j=\infty} \rho_3(F_i, F_j) = 0.$$

It follows from (8.5) and (8.3) that $W(F_i, d, x)$ converges, as $i \to \infty$, to a limit function $W(d, x)$ uniformly in d and x, i.e.,

$$(8.6) \qquad \lim_{i=\infty} W(F_i, d, x) = W(d, x)$$

uniformly in d and x. Hence

$$(8.7) \qquad \lim_{i=\infty} \int_D W(F_i, d, x) \, d\delta_x = \int_D W(d, x) \, d\delta_x$$

uniformly in x and δ. Because of (8.5), we have

$$(8.8) \qquad \lim_{i,j=\infty} \rho_1(F_i, F_j) = 0.$$

Hence there exists a distribution function $F_0(x)$ (not necessarily an element of Ω) such that

$$(8.9) \qquad \lim_{i=\infty} \rho_1(F_i, F_0) = 0.$$

It follows from (8.7) and (8.9) that

$$(8.10) \qquad \lim_{i \to \infty} \int_R \left[\int_D W(F_i, d, x) \, d\delta_x \right] dF_i(x) = \int_R \left[\int_D W(d, x) \, d\delta_x \right] dF_0(x)$$

uniformly in δ. Hence $\{F_i\}$ is a Cauchy sequence in the sense of the metric ρ_4 and Lemma 8.1 is proved.

Next we prove

LEMMA 8.2. *If D is conditionally compact in the sense of the metric (1.1) and if δ is any decision function, then for any $\epsilon > 0$ there exists a finite subset D^1 of D and a decision function δ^1 such that $\delta^1(D^1 \mid x) = 1$ identically in x and δ^1 is ϵ-equivalent to δ.*

PROOF. Since D is conditionally compact, it is possible to decompose D into a finite number of disjoint subsets $D_1, \cdots, D_u$ such that the diameter of D_j is less then ϵ $(j = 1, \cdots, u)$. Let d_j be an arbitrary but fixed point of $D_j (j = 1, \cdots, u)$ and let $\delta^1(x)$ be the decision function determined by the condition

$$(8.11) \qquad \delta^1(d_j \mid x) = \delta(D_j \mid x) \qquad\qquad (j = 1, \cdots, u).$$

Clearly

$$(8.12) \qquad \left| \int_D W(F, d, x) \, d\delta_x - \int_D W(F, d, x) \, d\delta_x^1 \right| \leq \epsilon$$

for all F and x. Hence,

$$(8.13) \qquad | r(F, \delta^1) - r(F, \delta) | \leq \epsilon$$

for all F and our lemma is proved.

We are now in a position to prove the main theorem.

THEOREM 8.1. *If the elements $F(x)$ of Ω are atomless, if Ω is conditionally compact in the sense of the metrics ρ_1 and ρ_2, and if D is conditionally compact in the the sense of the metric (1.1), then for any $\epsilon > 0$ and for any decision function $\delta(x)$ there exists an ϵ-equivalent nonrandomized decision function $\delta^*(x)$.*

PROOF. Because of Lemma 8.2, it is sufficient to prove our theorem for finite D. Thus, we shall assume that D consists of the elements $d_1, \cdots, d_m$. It is easy to verify that conditional compactness of Ω in the sense of both metrics ρ_1 and ρ_2 implies conditional compactness in the sense of the metric ρ_3, and because of Lemma 8.1, also in the sense of the metric ρ_4. Thus, conditional compactness of Ω in the sense of the metrics ρ_1 and ρ_2 implies the existence of a finite subset $\Omega^* = \{F_1, \cdots, F_k\}$ of Ω such that Ω^* is $\epsilon/2$-dense in Ω in the sense of the metric ρ_4. Let δ^* be a nonrandomized decision function that is equivalent to δ if Ω is replaced by Ω^*. The existence of such a δ^* follows from Theorem 3.1. Since Ω^* is $\epsilon/2$-dense in Ω (in the sense of the metric ρ_4), we have

$$(8.14) \qquad | r(F, \delta^*) - r(F, \delta) | \leq \epsilon \quad \text{for all} \quad F \text{ in } \Omega$$

and our theorem is proved.

We shall now introduce some notions with the help of which we shall be able to strengthen Theorem 3.1. For any measurable subset S of R, let

$$(8.15) \qquad r(F, \delta \mid S) = \int_S \left[\int_D W(F, d, x) \, d\delta_x \right] dF(x).$$

We shall refer to the above expression as the contribution of the set S to the risk. For any S we shall consider the following four metrics in Ω:

$$(8.16) \qquad \rho_{1S}(F_1, F_2) = \operatorname*{Sup}_S \left| \int_{S^*} dF_1 - \int_{S^*} dF_2 \right|$$

where S^* is any measurable subset of S,

$$(8.17) \qquad \rho_{2S}(F_1, F_2) = \operatorname*{Sup}_{d, x \in S} \left| W(F_1, d, x) - W(F_2, d, x) \right|,$$

$$(8.18) \qquad \rho_{3S}(F_1, F_2) = \rho_{1S}(F_1, F_2) + \rho_{2S}(F_1, F_2),$$

$$(8.19) \qquad \rho_{4S}(F_1, F_2) = \operatorname*{Sup}_\delta \left| r(F_1, \delta \mid S) - r(F_2, \delta \mid S) \right|.$$

Finally let the metric $\rho_S(d_1, d_2)$ in D be defined by

$$(8.20) \qquad \rho_S(d_1, d_2) = \operatorname*{Sup}_{F, x \in S} \left| W(F, d_1, x) - W(F, d_2, x) \right|.$$

We shall now prove the following stronger theorem:

THEOREM 8.2. *Let all elements F of Ω be atomless. If there exists a decomposition of R into a sequence $\{R_i\}(i = 1, 2, \cdots,$ ad inf.) of disjoint subsets such that Ω is conditionally compact in the sense of the metrics ρ_{1R_i} and ρ_{2R_i} for each i, and such that D is conditionally compact in the sense of the metric ρ_{R_i} for each i, then for any $\epsilon > 0$ and for any decision function δ there exists an ϵ-equivalent nonrandomized decision function δ^*.*

PROOF. Let $\{R_i\}$ be a decomposition of R for which the conditions of our theorem are fulfilled. Let $\{\epsilon_i\}$ be a sequence of positive numbers such that $\sum_{i=1}^\infty \epsilon_i = \epsilon$. Let $\delta^1(x)$ be a decision function such that $\delta_1(x) = \delta(x)$ for any x not in R_1, $\delta^1(x)$ is nonrandomized over R_1 (for any x in R_1, $\delta^1(x)$ assigns the probability one to a single point d in D) and such that

$$(8.21) \qquad \left| r(F, \delta \mid R_1) - r(F, \delta^1 \mid R_1) \right| \leqq \epsilon_1 \quad \text{for all } F.$$

The existence of such a decision function δ^1 follows from Theorem 8.1 (replacing R by R_1). After $\delta^1, \cdots, \delta^{i-1}$ have been defined ($i \geqq 1$), let δ^i be a decision function such that δ^i is nonrandomized over R^i, $\delta^i(x) = \delta^{i-1}(x)$ for all x in $\bigcup_{j=1}^{i-1} R_j$, $\delta^i(x) = \delta(x)$ for all x in $R - \bigcup_{j=1}^i R_j$ and such that

$$(8.22) \qquad \left| r(F, \delta^i \mid R_i) - r(F, \delta \mid R_i) \right| \leqq \epsilon_i \quad \text{for all } F \text{ in } \Omega.$$

The existence of such a decision function δ^i follows again from Theorem 8.1. Clearly $\delta^i(x)$ converges to a limit $\delta^*(x)$, as $i \to \infty$. This limit decision function $\delta^*(x)$ is obviously nonrandomized and satisfies the conditon

$$(8.23) \qquad\qquad | r(F, \delta \mid R_i) - r(F, \delta^* \mid R_i) | \leq \epsilon_i$$

for all i and F. Theorem 8.2 is an immediate consequence of this.

The conditions of Theorem 8.2 will be fulfilled for a wide class of statistical decision problems. For example, this is true for the decision problems which satisfy the following six conditions:

CONDITION 1. *The sample space R is a finite dimensional Euclidean space. All elements $F(x)$ of Ω are absolutely continuous.*

CONDITION 2. *Ω admits a parametric representation, i.e., each element F of Ω is associated with a parametric point θ in a finite dimensional Euclidean space E.*

We shall denote the density function $p(x)$ corresponding to the parameter point θ by $p(x, \theta)$.

CONDITION 3. *The set of parameter points θ which correspond to all elements F of Ω is a closed subset of E.*

We shall call this set of all parameter points θ the parameter space. Since there is a one-to-one correspondence between the elements F of Ω and the points θ of the parameter space, there is no danger of confusion if we denote the parameter space also by Ω.

CONDITION 4. *The density function $p(x, \theta)$ is continuous in $\theta \in \Omega$ for every x.*

CONDITION 5. *The loss $W(\theta, d)$ when θ is true and the decision d is made does not depend on x. D is conditionally compact in the sense of the metric $\rho(d_1, d_2) = \mathrm{Sup} \mid W(\theta, d_1) - W(\theta, d_2) \mid$.*

CONDITION 6. *For any bounded subset M of R, we have $\lim\limits_{\{|\theta|\to\infty\atop \theta\in\Omega\}} \int_M p(x, \theta)\, dx = 0$.*

We shall now show that the conditions of Theorem 8.2 are fulfilled for any decision problem that satisfies Conditions 1–6. Let S_i be the sphere in R with center at the origin and radius i. Let $R_1 = S_1$ and $R_i = S_i - \bigcup\limits_{j=1}^{i-1} R_j (i = 1, 2, \cdots$, ad inf.). Condition 5 implies that D is conditionally compact in the sense of the metric ρ_{R_i} for all i. It follows from Condition 5 and Theorem 2.1 in [3] that Ω is conditionally compact in the sense of the metric $\rho(\theta_1, \theta_2) = \mathrm{Sup} \mid W(\theta_1, d) - W(\theta_2, d) \mid$. Hence Ω is conditionally compact in the sense of the metric ρ_{2R_i} for each i. It remains to be shown that Ω is conditionally compact in the sense of the metric ρ_{1R_i} for each i. For this purpose, consider any sequence $\{\theta_j\}(j = 1, 2, \cdots$, ad inf.) of parameter points. There are 2 cases possible: (a) $\{\theta_j\}$ admits a subsequence that converges in the Euclidean sense to a finite point θ_0; (b) $\lim\limits_{j\to\infty} \mid \theta_j \mid = \infty$. Let us consider first the case (a) and let $\{\theta_j'\}$ be a subsequence of $\{\theta_j\}$ which converges to a finite point θ_0. It then follows from Condition 4 and a theorem of Robbins [4] that $\{\theta_j'\}$ is a Cauchy subsequence

in the sense of the metric ρ_{1R_i} for each i. In case (b), Condition 6 implies that
the sequence $\{\theta_j\}$ is a Cauchy sequence in the sense of the metric ρ_{1R_i} for each i.
Thus, Ω is conditionally compact in the sense of the metric ρ_{1R_i}. This completes
the proof of our assertion that a decision problem that satisfies Conditions 1–6,
satisfies also the conditions of Theorem 8.2.

9. Application to the theory of games. Translation of the results of Section 2
into the language of the theory of games is immediate and we shall do this only
very briefly. The function $W(F_i, d_j, x)$ $(i = 1, \cdots, p; j = 1, \cdots, m; x \in R)$,
of Section 1 is now called the pay-off function of a zero-sum two-person game.
The game is played as follows: Player I selects one of the integers $1, \cdots, p$,
say i, without communicating his choice to player II. A random observation
$x \in R$ on a chance variable whose distribution function is F_i is obtained and
communicated to player II. The latter chooses one of the integers $1, \cdots, m$,
say j. The game now ends with the receipt by player I and player II of the
respective sums $W(F_i, d_j, x)$ and $-W(F_i, d_j, x)$. Randomized (mixed) and
nonrandomized (pure) strategies are defined in the same manner as the cor-
responding decision functions in Section 1. When the distribution functions
$F_i(x)(i = 1, \cdots, p)$ are all atomless the obvious analogues of Theorems 3.1
and 3.2 hold.

It should be remarked that the usual definition of randomized (mixed) strat-
egy is not as general as the one given above. In the usual definition player II
chooses, by a random mechanism independent of the random mechanism which
yields the point x, some one of a (usually finite) number of nonrandomized
(pure) strategies, and then plays the game according to the nonrandomized
strategy selected. In our definition (used in [3]) the random choice is allowed
to depend on x. Clearly our method of randomization includes the usual one as
a special case. The relation between the two methods of randomization will be
discussed by two of the authors in a forthcoming paper [7].

Suppose that the number of possible decisions is at most denumerable, and
that the decision procedure consists in choosing at random and in advance of
the observations, one of a finite number of nonrandomized decision functions.
The sample space can be divided into an at most denumerable number of sets
in each of which only a finite number of decisions is possible (the possible de-
cisions vary from set to set). In each set our results are applicable. Since the
number of sets is denumerable the resultant decision function is measurable.
We conclude: It follows from our results that if a decision procedure consists of
selecting with preassigned probabilities one of a finite number of nonrandomized
decision functions with the number of possible decisions at most denumerably
infinite, and if the possible distributions are finite in number and atomless, then
there exists an equivalent nonrandomized decision function. More general
results can be obtained for this case (where one chooses at random and in ad-
vance of the observations, one of a finite number of nonrandomized decision
functions). By application of the methods of Sections 4 and 8 the requirement

that the number of possible decisions be denumerable can be easily removed. The procedures are straightforward and we omit them.

REFERENCES

[1] A. DVORETZKY, A. WALD, AND J. WOLFOWITZ, "Elimination of randomization in certain problems of statistics and of the theory of games," *Proc. Nat. Acad. Sci.*, Vol. 36 (1950), pp. 256-260.

[2] A. LYAPUNOV, "Sur les fonctions-vecteurs complètement additives," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 4 (1940), pp. 465-478.

[3] A. WALD, *Statistical Decision Functions*, John Wiley & Sons, 1950.

[4] HERBERT ROBBINS, "Convergence of distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 72-75.

[5] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[6] A. DVORETZKY, A. WALD, AND J. WOLFOWITZ, "Relations among certain ranges of vector measures," *Pacific Journal of Mathematics* (1951).

[7] A. WALD AND J. WOLFOWITZ, "Two methods of randomization in statistics and the theory of games," *Annals of Mathematics*, to be published.

ON MINIMAX STATISTICAL DECISION PROCEDURES
AND THEIR ADMISSIBILITY[1]

By Colin R. Blyth

University of California, Berkeley

Summary. This paper is concerned with the problem of making a decision on the basis of a sequence of observations on a random variable. Two loss functions, each depending on the distribution of the random variable, the number of observations taken, and the decision made, are assumed given. Minimax problems can be stated for weighted sums of the two loss functions, or for either one subject to an upper bound on the expectation of the other. Under suitable conditions it is shown that solutions of the first type of problem provide solutions for all problems of the latter types, and that admissibility for a problem of the first type implies admissibility for problems of the latter types. Two examples are given: Estimation of the mean of a random variable which is (1) normal with known variance, (2) rectangular with known range. The resulting minimax estimates are, with a small class of exceptions, proved admissible among the class of all procedures with continuous risk functions. The two loss functions are in each case the number of observations, and an arbitrary nondecreasing function of the absolute error of estimate. Extensions to a function of the number of observations for the first loss function are indicated, and two examples are given for the normal case where the sample size can or must be randomised among more than a consecutive pair of integers.

1. Introduction. We will consider a sequence X_1, X_2, X_3, $\cdots$ of independent random variables, each having the same distribution F, which belongs to a class Ω of possible distributions. A sequential decision procedure S is a rule by which, having observed x_1, $\cdots$, $x_m (m = 0, 1, 2, \cdots)$ we make one of the following decisions:

(a) Take an observation on X_{m+1}.

(b) Stop experimentation and make a terminal decision $d = d(x_1, \cdots, x_m)$. We will consider two non-negative loss functions $W_1(n, d, F)$ and $W_2(n, d, F)$. Each can be thought of as an economic loss incurred when the X's have distribution F and the terminal decision d is made after n observations have been taken. In the simplest applications one W will be a function of n only (cost of experimentation) and the other W will be a function of d and F only (loss incurred by making the decision d when the X's have distribution F). We will denote by $E(W_i \mid F, S)$ the expected value of W_i when the X's have distribution F and the decision procedure S is used. Let ξ be any probability measure defined on some class of subsets of Ω. We will denote by $E(W_i \mid \xi, S)$ the expected value

---

of W_i, given the (*a priori*) distribution ξ over Ω, when the decision procedure S is used.

Minimax problems, first considered by Wald, have been formulated in three ways for the situation just described. We may seek a decision procedure S which will (i) subject to an upper bound on $E(W_1 \mid F, S)$, minimise $\sup_{\Omega} E(W_2 \mid F, S)$; or (ii) subject to an upper bound on $E(W_2 \mid F, S)$, minimise $\sup_{\Omega} E(W_1 \mid F, S)$; or (iii) minimise $\sup_{\Omega} \{c\, E(W_1 \mid F, S) + E(W_2 \mid F, S)\}$, where $0 < c < \infty$, c constant. We will show that in certain cases it suffices to find solutions for all problems (iii) since these solutions provide solutions for all problems (i) and (ii).

We will first discuss the corresponding Bayes problems, not for their own interest, but because they can often be used to find solutions for the minimax problems in which we are really interested.

2. Bayes problems. The following three classes of Bayes problems have been considered: Given *a priori* the distribution ξ over Ω, we want to find a (Bayes) procedure which will

(i)′   subject to $E(W_1 \mid \xi, S) \leq L_1$,   minimise $E(W_2 \mid \xi, S)$,

(ii)′   subject to $E(W_2 \mid \xi, S) \leq L_2$,   minimise $E(W_1 \mid \xi, S)$,

(iii)′   minimise $r_c(\xi, S) = c E(W_1 \mid \xi, S) + E(W_2 \mid \xi, S)$.

Let $\mathcal{S}_c$ be the class of all solutions of problem (iii)′ for a given c, $0 < c < \infty$. Let $\mathcal{S} = \bigcup_{0 < c < \infty} \mathcal{S}_c$ be the class of all solutions of problems (iii)′, $0 < c < \infty$.

LEMMA 1. *If $S' \in \mathcal{S}$ has $E(W_1 \mid \xi, S') = L_1$, then S' is a solution of the problem* (i)′ *for this L_1. If S'' is any other solution of this problem* (i)′, *then $E(W_1 \mid \xi, S'') = L_1$ and $S'' \in \mathcal{S}$. Similarly for problems* (ii)′.

PROOF. Let $S' \in \mathcal{S}_c$. Suppose there exists a procedure S^* having

$$E(W_1 \mid \xi, S^*) \leq E(W_1 \mid \xi, S') = L_1,$$

$$E(W_2 \mid \xi, S^*) < E(W_2 \mid \xi, S').$$

Then

$$cE(W_1 \mid \xi, S^*) + E(W_2 \mid \xi, S^*) < cE(W_1 \mid \xi, S') + E(W_2 \mid \xi, S').$$

This implies $S' \notin \mathcal{S}_c$, which is false. This contradiction shows that no such S^* can exist. Hence S' is a solution of this problem (i)′.

If S'' is any other solution of this problem (i)′ we must have $E(W_2 \mid \xi, S'') = E(W_2 \mid \xi, S')$. Suppose that $E(W_1 \mid \xi, S'') < E(W_1 \mid \xi, S') = L_1$. Then

$$cE(W_1 \mid \xi, S'') + E(W_2 \mid \xi, S'') < cE(W_1 \mid \xi, S') + E(W_2 \mid \xi, S'),$$

implying the contradiction $S' \notin \mathcal{S}_c$. Hence $E(W_1 \mid \xi, S'') = E(W_1 \mid \xi, S') = L_1$. We therefore have $r_c(\xi, S'') = r_c(\xi, S')$, and so $S'' \in \mathcal{S}_c$.

LEMMA 2. *If $S' \epsilon \, \mathcal{S}$, $S'' \epsilon \, \mathcal{S}$, then*

$$E(W_1 \mid \xi, S') < E(W_1 \mid \xi, S'') \longleftrightarrow E(W_2 \mid \xi, S') > E(W_2 \mid \xi, S''),$$

$$E(W_1 \mid \xi, S') = E(W_1 \mid \xi, S'') \longleftrightarrow E(W_2 \mid \xi, S') = E(W_2 \mid \xi, S'').$$

LEMMA 3. *If $S' \epsilon \, \mathcal{S}_{c'}$, and $S'' \epsilon \, \mathcal{S}_{c''}$, where $c' < c''$, then*

$$E(W_1 \mid \xi, S') \geqq E(W_1 \mid \xi, S''),$$

$$E(W_2 \mid \xi, S') \leqq E(W_2 \mid \xi, S'').$$

PROOF. Assume one of the following:

$$L_1 = E(W_1 \mid \xi, S') < E(W_1 \mid \xi, S'') = L_1 + r,$$

$$L_2 = E(W_2 \mid \xi, S') > E(W_2 \mid \xi, S'') = L_2 - s.$$

The other then follows from Lemma 2. Write $c'' = c' + a$. Here $r > 0, s > 0, a > 0$. Then

$$r_{c'}(\xi, S') = c'L_1 + L_2,$$

$$r_{c'}(\xi, S'') = c'L_1 + L_2 + (c'r - s),$$

$$r_{c''}(\xi, S') = c'L_1 + L_2 + aL_1,$$

$$r_{c''}(\xi, S'') = c'L_1 + L_2 + aL_1 + (c'r - s + ar).$$

Now $$S' \epsilon \, \mathcal{S}_{c'} \rightarrow c'r - s \geqq 0,$$

and $$S'' \epsilon \, \mathcal{S}_{c''} \rightarrow c'r - s + ar \leqq 0.$$

Since $ar > 0$ these last two results cannot both be true. This contradiction shows that neither of the assumed inequalities can be true, and proves the lemma.

Let us write

$$\underline{L}_1 = \inf_{s \epsilon \mathcal{S}} E(W_1 \mid \xi, S), \qquad \overline{L}_1 = \sup_{s \epsilon \mathcal{S}} E(W_1 \mid \xi, S),$$

$$\underline{L}_2 = \inf_{s \epsilon \mathcal{S}} E(W_2 \mid \xi, S), \qquad \overline{L}_2 = \sup_{s \epsilon \mathcal{S}} E(W_2 \mid \xi, S),$$

where the improper value ∞ is admitted for the upper bounds.

LEMMA 4.

$$E(W_1 \mid \xi, S) < \underline{L}_1 \rightarrow E(W_2 \mid \xi, S) = \infty,$$

$$E(W_2 \mid \xi, S) < \underline{L}_2 \rightarrow E(W_1 \mid \xi, S) = \infty.$$

PROOF. Suppose that S is a procedure for which $E(W_1 \mid \xi, S) = L_1 < \underline{L}_1$ and $E(W_2 \mid \xi, S) = L_2 < \infty$.

If $\overline{L}_2 = \infty$, there exists some $S_c \, \epsilon \, \mathcal{S}_c$ having $E(W_1 \mid \xi, S_c) \geqq \underline{L}_1$ and $E(W_2 \mid \xi, S_c) > L_2$; but we would then have $r_c(\xi, S_c) > r_c(\xi, S)$, contradicting the fact that $S_c \, \epsilon \, \mathcal{S}_c$.

If $\bar{L}_2 < \infty$, then for $S_c \, \epsilon \, \mathcal{S}_c$ we have

$$cE(W_1 \mid \xi, S_c) + E(W_2 \mid \xi, S_c) \geqq cL_1 + L_2 > cL_1 + L_2$$

for c sufficiently large, again contradicting the fact that $S_c \, \epsilon \, \mathcal{S}_c$. This completes proof of the first part of the lemma; the second part is proved in the same way.

Lemma 4 shows that no problem (i)′ with $L_1 < \underline{L}_1$ has a solution. Lemmas 2 and 4 show that if $S \, \epsilon \, \mathcal{S}$ has $E(W_1 \mid \xi, S) = \bar{L}_1$, then $E(W_2 \mid \xi, S) = \underline{L}_2$ and S is a solution of all problems (i)′ with $L_1 \geqq \bar{L}_1$. Similar remarks hold for problems (ii)′.

THEOREM. *If for every L_1 satisfying $\underline{L}_1 \leqq L_1 \leqq \bar{L}_1$, there exists $S \, \epsilon \, \mathcal{S}$ having $E(W_1 \mid \xi, S) = L_1$, then the class of all solutions of problems (i)′ with $\underline{L}_1 \leqq L_1 \leqq \bar{L}_1$ coincides with $\mathcal{S}$. Similarly for problems (ii)′. If $\bar{L}_1 = \infty$ or $\bar{L}_2 = \infty$ the appropriate equality signs must be omitted.*

This theorem is an immediate consequence of Lemma 1.

NOTE. From monotonicity (Lemma 3) we know that as $c \to c^0$ from one side and $S_c \, \epsilon \, \mathcal{S}_c$, $E(W_1 \mid \xi, S_c) \to$ some limit L_1 from one side and $E(W_2 \mid \xi, S_c) \to$ some limit L_2 from one side. If this implies the existence of a procedure S having $E(W_1 \mid \xi, S) = L_1$ and $E(W_2 \mid \xi, S) = L_2$ whenever L_1 and L_2 are finite, it is easy to show that $S \, \epsilon \, \mathcal{S}_{c^0}$, and that the conditions for the theorem are satisfied. However, the conditions themselves are usually easy to check once we have found $\mathcal{S}$.

Suppose that for a given Ω, ξ, W_1, W_2 we have found the class $\mathcal{S}$ of all solutions of problems (iii)′, $0 < c < \infty$, and find the conditions for the above theorem satisfied. Solving any problem (i)′ or (ii)′ is now reduced to choosing the appropriate member of $\mathcal{S}$.

3. Minimax problems. The following three classes of minimax problems have been considered: We want to find a (minimax) procedure which will

(i)   subject to $\sup_\Omega E(W_1 \mid F, S) \leqq L_1$,   minimise $\sup_\Omega E(W_2 \mid F, S)$,

(ii)   subject to $\sup_\Omega E(W_2 \mid F, S) \leqq L_2$,   minimise $\sup_\Omega E(W_1 \mid F, S)$,

(iii)   minimise $\sup_\Omega \{cE(W_1 \mid F, S) + E(W_2 \mid F, S)\}$.

If there is an *a priori* distribution ξ which is least favorable in problem (iii)′ for all c, $0 < c < \infty$, this distribution is also least favorable for all problems (i)′ and (ii)′. The Bayes solutions with respect to this distribution are minimax solutions of the corresponding problems stated in this section. In many problems, however, this easy approach is not available.

LEMMA 5. *Suppose some problem* (iii) *has a solution S' with*

$$\sup_\Omega E(W_1 \mid F, S') = L_1, \qquad \sup_\Omega E(W_2 \mid F, S') = L_2,$$

$$\sup_\Omega \{cE(W_1 \mid F, S') + E(W_2 \mid F, S')\} = cL_1 + L_2.$$

(*These conditions will in particular hold if either* $\sup_{\Omega} E(W_1 \mid F, S') = L_1$ *and* $E(W_2 \mid F, S') \equiv L_2$, *or* $\sup_{\Omega} E(W_2 \mid F, S') = L_2$ *and* $E(W_1 \mid F, S') \equiv L_1$.) *Then* S' *is a solution of the problem* (i) *with this* L_1, *and a solution of the problem* (ii) *with this* L_2.

PROOF. Suppose there is a procedure S having

$$\sup_{\Omega} E(W_1 \mid F, S) \leqq L_1, \quad \sup_{\Omega} E(W_2 \mid F, S) < L_2.$$

Then we would have

$$\sup_{\Omega} \{cE(W_1 \mid F, S) + E(W_2 \mid F, S)\} \leqq c \sup_{\Omega} E(W_1 \mid F, S)$$

$$+ \sup_{\Omega} E(W_2 \mid F, S) < cL_1 + L_2 = \sup_{\Omega} \{cE(W_1 \mid F, S') + E(W_2 \mid F, S')\},$$

contradicting the fact that S' is a solution of some problem (iii). Hence no such S can exist, and S' is a solution of the problem (i) with this L_1. Similarly S' is a solution of the problem (ii) with this L_2.

Let $\mathcal{C}$ be any class of solutions of problems (iii), each member S of which satisfies the condition

$$\sup_{\Omega} \{E(W_1 \mid F, S) + E(W_2 \mid F, S)\} = \sup_{\Omega} E(W_1 \mid F, S) + \sup_{\Omega} E(W_2 \mid F, S).$$

Let $\mathcal{C}_c$ denote those members of $\mathcal{C}$ which are solutions of the problem (iii) for this particular c. Then the following two lemmas can be proved in exactly the same way as the corresponding lemmas of Section 2.

LEMMA 2a. *If* $S' \epsilon \mathcal{C}$, $S'' \epsilon \mathcal{C}$, *then*

$$\sup_{\Omega} E(W_1 \mid F, S') < \sup_{\Omega} E(W_1 \mid F, S'') \longleftrightarrow \sup_{\Omega} E(W_2 \mid F, S') > \sup_{\Omega} E(W_2 \mid F, S''),$$

and

$$\sup_{\Omega} E(W_1 \mid F, S') = \sup_{\Omega} E(W_1 \mid F, S'') \longleftrightarrow \sup_{\Omega} E(W_2 \mid F, S') = \sup_{\Omega} E(W_2 \mid F, S'').$$

LEMMA 3a. *If* $S' \epsilon \mathcal{C}_{c'}$ *and* $S'' \epsilon \mathcal{C}_{c''}$, *where* $c' < c''$, *then*

$$\sup_{\Omega} E(W_1 \mid F, S') \geqq \sup_{\Omega} E(W_1 \mid F, S'')$$

and

$$\sup_{\Omega} E(W_2 \mid F, S') \leqq \sup_{\Omega} E(W_2 \mid F, S'').$$

Suppose that we have found such a class $\mathcal{C}$ of solutions of problems (iii) and that there exists $S \epsilon \mathcal{C}$ having $\sup_{\Omega} E(W_i \mid F, S) = L_i$ whenever $\inf_{n,d,F} W_i(n, d, F) \leqq L_i \leqq \sup_{n,d,F} W_i(n, d, F)$, $i = 1, 2$. (Omit appropriate equality signs if either upper bound is ∞). Then solving any problem (i) or (ii) is reduced to choosing the appropriate member of $\mathcal{C}$.

In order to find solutions of problems (iii) in the examples we consider, the following lemma, which is due to E. Lehmann, will be needed.

LEMMA 6. *Consider the minimax problem of finding a procedure which minimises* $\sup_{\Omega} r(F, S)$. *(This may be subject to conditions as in* (i) *and* (ii), *or not as in* (iii).) *Let S_k be a Bayes procedure with respect to the a priori distribution ξ_k over Ω,* $k = 1, 2, \cdots$. *Then for any procedure S,*

$$\sup_{\Omega} r(F, S) \geq r(\xi_k, S) \geq r(\xi_k, S_k)$$

for all k. Therefore

$$\sup_{\Omega} r(F, S) \geq \lim_{k \to \infty} \sup r(\xi_k, S_k).$$

A sufficient condition for the procedure S_0 to be minimax is therefore

$$r(F, S_0) \leq \lim_{k \to \infty} \sup r(\xi_k, S_k)$$

for all $F \epsilon \Omega$.

4. Admissibility. Admissible procedures (not necessarily solutions) for the problems stated in Section 3 are defined as follows:

A procedure S is admissible for a particular problem (iii) if there is no procedure S^* having

$$r_e(F, S^*) \leq r_e(F, S) \quad \text{for all} \quad F \epsilon \Omega,$$

with strict inequality for at least one $F \epsilon \Omega$, where $r_e(F, S) = cE(W_1 \mid F, S) + E(W_2 \mid F, S)$.

A procedure S is admissible for a particular problem (i) if there is no procedure S^* having

$$\sup_{\Omega} E(W_1 \mid F, S^*) \leq L_1,$$

and

$$E(W_2 \mid F, S^*) \leq E(W_2 \mid F, S) \quad \text{for all} \quad F \epsilon \Omega,$$

with strict inequality for at least one $F \epsilon \Omega$. Admissibility is defined in a similar way for problem (ii).

LEMMA 7. *Suppose S is an admissible procedure for some problem* (iii). *Then if $E(W_1 \mid F, S) \underset{\Omega}{\equiv} L_1$, S is admissible for the problem* (i) *with this L_1. And if* $E(W_2 \mid F, S) \underset{\Omega}{\equiv} L_2$, S *is admissible for the problem* (ii) *with this L_2.*

PROOF. Suppose that $E(W_1 \mid F, S) \underset{\Omega}{\equiv} L_1$ and that S is not admissible for the problem (i) with this L_1. Then there is a procedure S^* having $\sup_{\Omega}$

$E(W_1 \mid F, S^*) \leqq L_1$; and $E(W_2 \mid F, S^*) \leqq E(W_2 \mid F, S)$ for all $F \in \Omega$, with strict inequality for at least one $F \in \Omega$. We therefore have

$$r_c(F, S^*) = cE(W_1 \mid F, S^*) + E(W_2 \mid F, S^*)$$
$$\leqq cL_1 + E(W_2 \mid F, S) = cE(W_1 \mid F, S) + E(W_2 \mid F, S) = r_c(F, S)$$

for all $F \in \Omega$, with strict inequality for at least one $F \in \Omega$. That is, S cannot be admissible for any problem (iii), a contradiction which proves the first part of the lemma. The second part is proved in the same way.

If for a problem there is a least favorable distribution for which the Bayes solution is unique, this is the unique minimax solution and is therefore admissible. If Ω is a parametric family and all possible procedures have risks continuous in the parameter θ, and λ is a least favorable distribution which assigns positive probability to every interval of values of θ, then any Bayes solution for this λ is minimax and admissible. When can we conclude that minimax solutions obtained by the method of Lemma 6 are admissible? In Sections 5 and 7 we will show for particular examples that the solutions so obtained, except for trivial exceptions, are all admissible among the class of procedures with continuous risk functions. We might hope that all constant risk minimax solutions so obtained are admissible, but will see that this is not so.

The method used here for proving admissibility of minimax solutions involves examination of the Bayes solutions used to obtain them. In the examples considered, if W_2 is continuous, this method works both for classical fixed sample size problems and for the sequential problems (i), (ii), (iii) subject to the additional restriction that the number of observations is bounded.

Admissibility is proved for a number of examples by Hodges and Lehmann in [4] by a completely different method, which involves no appeal to Bayes solutions, and which works for certain fixed sample size problems in which the method of this paper fails. Their method, however, is restricted to number of observations and squared error of estimate for loss functions, and among sequential problems will handle only (i), again subject to the additional restriction that the number of observations is bounded.

5. Example: normal. Let $X_1, X_2, \cdots$ be a sequence of independent random variables, each being $N(\theta, 1)$, i.e., normal with mean θ and variance 1. A point estimate z is wanted for the mean θ. Let

$$W_1(n, z, \theta) = n, \qquad W_2(n, z, \theta) = W(z - \theta),$$

where W is a non-decreasing function of $\mid z - \theta \mid$. The three classes of minimax problems are

(i)   subject to $\sup_\theta E_\theta(n) \leqq M$,   minimise $\sup_\theta E_\theta W(z - \theta)$,

(ii)   subject to $\sup_\theta E_\theta W(z - \theta) \leqq L$,   minimise $\sup_\theta E_\theta(n)$,

(iii)   minimise $\sup_\theta \{cE_\theta(n) + E_\theta W(z - \theta)\}$.

NOTE. This problem was first considered by Stein and Wald in [1]. They solved problems (i) and (ii) for the case $W(z - \theta) = 0$ or 1 according as $|z - \theta| \leqq a/2$ or $> a/2$; their estimates are thus confidence intervals of fixed length a. For this same case Wolfowitz in [2] solved problems (iii) and showed that these solutions provide solutions for problems (i) and (ii). Wolfowitz also obtained solutions of problems (iii) for the general $W(z - \theta)$, non-decreasing in $|z - \theta|$. The question of admissibility is not considered in [1] or [2].

The remainder of this section will be concerned with proving the following results.

THEOREM. *To a given c there corresponds either an integer N or a pair of consecutive integers N, $N + 1$. A class of solutions of the problem* (iii) *for this c are procedures in which the only possible sample sizes are N (or N, N + 1) and in which the estimate used is $\frac{1}{n} \sum_{i=1}^{n} X_i$ if $n > 0$. If $N \neq 0$, all such solutions are admissible among the class of procedures with continuous risk functions. The class of solutions so obtained, $0 < c < \infty$, provides solutions for all problems* (i) *and* (ii).

We will find solutions for problems (iii) by first finding Bayes solutions for the corresponding problems (iii)$'$ when θ has the a priori distribution $N(0, \sigma^2)$. The Bayes problem is to find a sequential estimation procedure which will minimise the risk

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \{cE_\theta(n) + E_\theta W(z - \theta)\} e^{-(1/2\sigma^2)\theta^2} \, d\theta.$$

We will assume that $W(z - \theta)$ increases with $|z - \theta|$ slowly enough so that

$$\int_{-\infty}^{\infty} E_\theta W(z \stackrel{l}{-} \theta) e^{-(1/2\sigma^2)(\theta - \mu)^2} \, d\theta < \infty$$

for some σ_0, μ_0, z_0, and hence for all $\sigma < \sigma_0$, μ, z.

Let us first determine what should be our estimate z for θ if we stop after having observed $x_1, \cdots, x_m$. For this we need to know the a posteriori distribution

$$p(\theta \mid x_1, \cdots, x_m) = p(\theta, x_1, \cdots, x_m)/p(x_1, \cdots, x_m)$$
$$= c_1(x_1, \cdots, x_m) e^{-(1/2\sigma^2)\theta^2} e^{-\frac{1}{2}\Sigma_1^m (x_i - \theta)^2}$$
$$= c_2(x_1, \cdots, x_m) e^{-((m\sigma^2+1)/2\sigma^2)(\theta - (\sigma^2/(m\sigma^2+1))\Sigma_1^m x_i)^2}.$$

That is, θ, given $x_1, \cdots, x_m$, is $N\left(\dfrac{\sigma^2}{m\sigma^2 + 1} \sum_1^m x_i, \dfrac{\sigma^2}{m\sigma^2 + 1}\right)$. Given that we observe $x_1, \cdots, x_m$ and then stop and estimate $z(x_1, \cdots, x_m)$ for θ, our (a posteriori) risk is therefore

$$cm + \frac{\sqrt{m\sigma^2+1}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} W(z - \theta) e^{-((m\sigma^2+1)/2\sigma^2)(\theta - (\sigma^2/(m\sigma^2+1))\Sigma_1^m x_i)^2} \, d\theta.$$

Since $W(z - \theta)$ is a non-decreasing function of $|z - \theta|$, this risk is clearly minimised by choosing $z = \dfrac{\sigma^2}{m\sigma^2 + 1} \sum_1^m x_i$, where we interpret $\sum_1^m x_i = 0$ if $m = 0$. The minimum value is

$$r_{c,\sigma}(m) = cm + \frac{\sqrt{m\sigma^2 + 1}}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} W(y)e^{-((m\sigma^2+1)/2\sigma^2)y^2}\, dy.$$

This does not depend on the observations, but only on the number of observations. Since $r_{c,\sigma} \to \infty$ as $m \to \infty$ it is clear that the sequence $r_{c,\sigma}(m): m = 0,$ $1, 2, \cdots$ assumes a minimum value at a finite set $n_1', \cdots, n_{p'}'[p' = p'(c, \sigma)]$ of integers m. Hence if θ is $N(0, \sigma^2)$ *a priori*, any of the following procedures is Bayes: The only possible sample sizes are $n_1', \cdots, n_{p'}'$; if the sample size is m, the estimate $z = \dfrac{\sigma^2}{m\sigma^2 + 1} \sum_1^m x_i$ is used for θ.

To obtain minimax procedures, consider a sequence of σ's tending to ∞. As $\sigma \to \infty$,

$$r_{c,\sigma}(m) \to r_c(m) = cm + \sqrt{\frac{m}{2\pi}} \int_{-\infty}^{\infty} W(y)e^{-(m/2)y^2}\, dy$$

for $m = 1, 2, \cdots$, and $r_{c,\sigma}(0) \to r_c(0) = \sup_y W(y)$.

Clearly $r_c(m): m = 0, 1, 2, \cdots$ assumes a minimum value at a finite set $n_1, \cdots, n_p [p = p(c)]$ of integers m.

Consider the following class $\mathcal{C}_c'$ of sequential procedures: The only possible sample sizes are $n_1, \cdots, n_p$. If the sample size is 0, estimate 0 for θ (any estimate whatever will do as well). If the sample size is $m > 0$ estimate $z = \dfrac{1}{m} \sum_1^m x_i$ for θ. Writing $n_1 < n_2 < \cdots < n_p$, the risk of any such procedure, if $n_1 = 0$, is

$$r_c^*(\theta) = P(n = 0)W(\theta) + \sum_{i=2}^{p} \left\{ P_\theta(n = n_i)\left[cn_i + E_\theta W\left(\frac{1}{n_i}\sum_1^{n_i} x_j - \theta\right)\right]\right\}$$

$$= P(n = 0)W(\theta) + \sum_{i=2}^{p} \left\{ P_\theta(n = n_i)\left[cn_i + \sqrt{\frac{n_i}{2\pi}} \int_{-\infty}^{\infty} W(y)e^{-(m/2)y^2}\, dy\right]\right\}$$

$$\leq P(n = 0) \sup_y W(y) + \sum_{i=2}^{p} P_\theta(n = n_i)r_c(n_i)$$

$$= \sup_y W(y) = r_c(n_i), \qquad\qquad i = 2, \cdots, p, \quad \text{for all} \quad \theta.$$

Similarly, if $n_1 \neq 0$, it is easy to show that

$$r_c^*(\theta) = r_c(n_i), \qquad i = 1, \cdots, p, \text{ for all } \theta.$$

It follows at once from Lemma 6 that **every member of $\mathcal{C}_c'$ is a minimax procedure** for the problem (iii) with this c.

We will next show that

$$r_c(m) = cm + \sqrt{\frac{m}{2\pi}} \int_{-\infty}^{\infty} W(y) e^{-(m/2)y^2}\, dy \quad \text{for} \quad m > 0,$$

$$= \sup_y W(y) \quad \text{for} \quad m = 0$$

is a convex function of m. Let m_0 be the smallest integer for which $r_c(m) < \infty$; this is the same for all c. Then $r_c(m)$ is continuous in m for all $m \geq m_0$. Convexity of $r_c(m)$ is equivalent to convexity of

$$g(m) = \sqrt{m} \int_{-\infty}^{\infty} W(y) e^{-(m/2)y^2}\, dy.$$

It is easy to show that for $m_0 \leq m < \infty$, differentiation under the integral sign any number of times is valid for $g(m)$. Therefore

$$g'(m) = \frac{1}{2\sqrt{m}} \int_0^{\infty} W(y) e^{-(m/2)y^2}(1 - my^2)\, dy,$$

$$g''(m) = \frac{1}{4m\sqrt{m}} \int_0^{\infty} W(y) e^{-(m/2)y^2}(m^2 y^4 - 2my^2 - 1)\, dy$$

$$= \frac{1}{4m^2} \int_0^{\infty} W\left(\frac{x}{\sqrt{m}}\right) e^{-\frac{1}{2}x^2}(x^4 - 2x^2 - 1)\, dx.$$

Now

$$x^4 - 2x^2 - 1 < 0 \quad \text{for} \quad 0 \leq x < \sqrt{1 + \sqrt{2}},$$

$$x^4 - 2x^2 - 1 > 0 \quad \text{for} \quad \sqrt{1 + \sqrt{2}} < x.$$

Also, $W(y)$ is non-decreasing as $y > 0$ increases and we will exclude from consideration the trivial case $W(y) \equiv$ constant. It follows that

$$g''(m) > \frac{1}{4m^2} \int_0^{\sqrt{1+\sqrt{2}}} W\left(\sqrt{\frac{1 + \sqrt{2}}{m}}\right) e^{-\frac{1}{2}x^2}(x^4 - 2x^2 - 1)\, dx$$

$$+ \frac{1}{4m^2} \int_{\sqrt{1+\sqrt{2}}}^{\infty} W\left(\sqrt{\frac{1 + \sqrt{2}}{m}}\right) e^{-\frac{1}{2}x^2}(x^4 - 2x^2 - 1)\, dx$$

$$= \frac{1}{4m^2} W\left(\sqrt{\frac{1 + \sqrt{2}}{m}}\right) \int_0^{\infty} e^{-\frac{1}{2}x^2}(x^4 - 2x^2 - 1)\, dx = 0.$$

That is, $g(m)$ is strictly convex for all $m \geq m_0$. Hence $r_c(m)$ is continuous and strictly convex for $m \geq m_0$.

For any given c, it follows that $r_c(m): m = 0, 1, 2, \cdots$ is smallest for at most two consecutive integers m. If at two consecutive integers, these must be on opposite sides of the m which minimises $r_c(m)$. (Thus $p = 1$ or 2. The same results are now obvious for any $r_{c,\sigma}(m)$, given c, σ.)

For all c sufficiently large, $r_c(m): m = 0, 1, 2, \cdots$ is minimised by $m = m_0$

only. Now, for any given m, $r_c(m)$ and $\partial r_c(m)/\partial m$ and $r_c(m + 1) - r_c(m)$ are continuous and strictly increasing functions of c, $0 < c < \infty$. Therefore as we decrease c continuously from such a sufficiently large value, $r_c(m):m = 0, 1, 2, \cdots$ remains smallest for $m = m_0$ only, until a point c^1 is reached for which $r_{c^1}(m):m = 0, 1, 2, \cdots$ is minimised by $m = m_0$ and $m = m_0 + 1$. As we continue to decrease c, for ε sufficiently small and $c^1 - \varepsilon < c < c^1$, $r_c(m):m = 0, 1, 2, \cdots$ is clearly smallest for $m = m_0 + 1$ only. This remains true until we reach a point c^2 for which $r_{c^2}(m):m = 0, 1, 2, \cdots$ is minimised by $m = m_0 + 1$ and $m = m_0 + 2$. As we continue to decrease c, $r_c(m):m = 0, 1, 2, \cdots$ is smallest for larger and larger m's, which tend to ∞ as $c \to 0$, because, for a given m, $\partial r_c(m)/\partial m < 0$ for all c sufficiently small. We note that only for a denumerable set of c's is $r_c(m):m = 0, 1, 2, \cdots$ minimised by two consecutive m's; for almost all c's this minimum occurs for only one m.

Let $\mathcal{C}_c$ consist of those members of $\mathcal{C}_c'$ in which the sample size does not depend on θ. Included are the procedures in which the sample size is randomised, independently of the observations, between the possible sample sizes. Let $\mathcal{C} = \bigcup_{0 < c < \infty} \mathcal{C}_c$. Now $E_\theta(n)$ is constant for any member of $\mathcal{C}$, implying $\sup_\theta \{E_\theta(n) + E_\theta(W)\} = \sup_\theta E_\theta(n) + \sup_\theta E_\theta(W)$. Lemmas 5, 2a and 3a are therefore valid for $\mathcal{C}$.

For every M, $m_0 \leq M < \infty$ there is clearly a member of $\mathcal{C}$ having $E_\theta(n) \equiv M$. Using continuity considerations it is easy to show that for every L, $0 < L < \infty$, there is a member of $\mathcal{C}$ having $\sup_\theta E_\theta(W) = L$. It follows from Lemma 5 that $\mathcal{C}$ contains a solution for every problem (i) with $M \geq m_0$ (problems (i) with $M < m_0$ have no solutions) and a solution for every problem (ii). Selection of the appropriate member of $\mathcal{C}$ is obvious for any particular problem (i), requires successive approximation for any particular problem (ii).

Are the members of $\mathcal{C}' = \bigcup_{0 < c < \infty} \mathcal{C}_c'$ admissible for the problems (iii) for which they are solutions? We will answer this question first for those members of $\mathcal{C}'$ for which 0 is not a possible sample size.

For a given c, suppose that $r_c(m):m = 0, 1, 2, \cdots$ is minimised by $m = N \neq 0$ only, or by $m = N \neq 0$ and $m = N + 1$ only. Since, for every m, $r_{c,\sigma}(m) \to r_c(m)$ as $\sigma \to \infty$, it is clear that if θ has the distribution $\lambda_\sigma = N(0, \sigma^2)$ a priori with σ sufficiently large, say $\sigma > K_1$, then N and $N + 1$ are the only possible sample sizes for a Bayes solution. We observe further that

$$r_{c,\sigma}(N) = cN + \frac{1}{\sqrt{2\pi}} \sqrt{N + \frac{1}{\sigma^2}} \int_{-\infty}^{\infty} W(y) e^{-((N+1/\sigma^2)/2)y^2} \, dy,$$

$$r_{c,\sigma}(N + 1) = c(N + 1) + \frac{1}{\sqrt{2\pi}} \sqrt{N + \frac{1}{\sigma^2} + 1} \int_{-\infty}^{\infty} W(y) e^{-((N+1/\sigma^2+1)/2)y^2} \, dy.$$

If $r_c(N) \leq r_c(N + 1)$, as we are assuming, it follows from the convexity of $g(m) = \sqrt{m} \int_0^\infty W(y) \, e^{-(m/2)y^2} \, dy$ that $r_{c,\sigma}(N) < r_{c,\sigma}(N + 1)$. Hence N is the only

possible sample size for a Bayes procedure, $\sigma > K_1$. Therefore, for this given c the (minimax) risk function for every member of $\mathcal{C}'_c$ is

$$r(\theta) \equiv r = cN + \frac{1}{\sqrt{2\pi}} \sqrt{N} \int_{-\infty}^{\infty} W(y)e^{-(N/2)y^2} \, dy,$$

and the Bayes risk for *a priori* λ_σ, $\sigma > K_1$, is

$$r_\sigma = cN + \frac{1}{\sqrt{2\pi}} \sqrt{N + \frac{1}{\sigma^2}} \int_{-\infty}^{\infty} W(y)e^{-((N+1/\sigma^2)/2)y^2} \, dy.$$

If the procedures in $\mathcal{C}'_c$ are non-admissible for this problem (iii) there must exist a procedure S^* having risk function $r^*(\theta) \leqq r$ for all θ, with strict inequality for at least one θ. Assuming $r^*(\theta)$ continuous this implies strict inequality for some interval of values of θ. We can therefore find two constants a and k, $0 < a < r$ and $0 < k < \infty$ such that

$$\frac{1}{2k} \int_{-k}^{k} r^*(\theta) \, d\theta = a.$$

Also, given any fixed ε, $0 < \varepsilon < 1 - a/r$, we can find $K > K_1$ so large that for $-k \leqq \theta \leqq k$,

$$1 - \varepsilon < e^{-(1/2\sigma^2)\theta^2} < 1 \quad \text{whenever} \quad \sigma > K.$$

Then for all $\sigma > K$ we have

$$\int_{-\infty}^{\infty} r^*(\theta)\lambda_\sigma(\theta) \, d\theta = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} r^*(\theta)e^{-(1/2\sigma^2)\theta^2} \, d\theta$$

$$\leqq \frac{1}{\sqrt{2\pi}\sigma} \int_{-k}^{k} r^*(\theta)e^{-(1/2\sigma^2)\theta^2} \, d\theta + \frac{2}{\sqrt{2\pi}\sigma} \int_{k}^{\infty} re^{-(1/2\sigma^2)\theta^2} \, d\theta$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-k}^{k} r^*(\theta)e^{-(1/2\sigma^2)\theta^2} \, d\theta + r - \frac{2r}{\sqrt{2\pi}\sigma} \int_{0}^{k} e^{-(1/2\sigma^2)\theta^2} \, d\theta$$

$$\leqq \frac{1}{\sqrt{2\pi}\sigma} \int_{-k}^{k} r^*(\theta) \cdot 1 \, d\theta + r - \frac{2r}{\sqrt{2\pi}\sigma} \int_{0}^{k} (1 - \varepsilon) \, d\theta$$

$$= \frac{1}{\sqrt{2\pi}\sigma} 2ka + r - \frac{2r}{\sqrt{2\pi}\sigma} k(1 - \varepsilon)$$

$$= r - \frac{b}{\sigma},$$

where $b = \dfrac{2k(r - a - r\varepsilon)}{\sqrt{2\pi}} > 0$ is a constant.

Now the Bayes risk for λ_σ, $\sigma > K$, is

$$r_\sigma = r - \frac{2}{\sqrt{2\pi}} \left\{ \sqrt{N} \int_{0}^{\infty} W(y)e^{-(N/2)y^2} \, dy \right.$$

$$\left. - \sqrt{N + \frac{1}{\sigma^2}} \int_{0}^{\infty} W(y)e^{-((N+1/\sigma^2)/2)y^2} \, dy \right\}.$$

We have seen that for $m \geq N$, the function $g(m) = \sqrt{m} \int_0^\infty W(y) \, e^{-(m/2)y^2} \, dy$ has continuous derivatives $g'(m) < 0$ and $g''(m) > 0$. It follows that

$$r_e \geq r + \frac{2}{\sqrt{2\pi}} \, g'(N) \frac{1}{\sigma^2},$$

$g'(N)$ being a negative constant. It is clear that for σ sufficiently large,

$$r_e \geq r + \frac{2}{\sqrt{2\pi}} \, g'(N) \frac{1}{\sigma^2} > r - \frac{b}{\sigma}$$

$$\geq \int_{-\infty}^\infty r^*(\theta) \lambda_\sigma(\theta) \, d\theta.$$

But this contradicts the fact that r_e is the Bayes risk for λ_σ, and so no such S^* can exist. Therefore, if 0 is not a possible sample size for members of $\mathcal{C}'_e$, every member of $\mathcal{C}'_e$ is admissible among the class of procedures with continuous risk functions, for the problem (iii) with this c.

Furthermore, $E_\theta(n)$ and $E_\theta(W)$ are both constants for members of $\mathcal{C}$ which belong to such a $\mathcal{C}'_e$. It follows from Lemma 7 that such members of $\mathcal{C}$ are admissible among the class of procedures with continuous risk functions, for the problems (i) and (ii) for which they are minimax.

If W is continuous and the number of observations is bounded, it can be shown that $r^*(\theta)$ is continuous. Thus if W is continuous, we have proved admissibility among the class of procedures with n bounded.

There remains the question of admissibility for those $\mathcal{C}'_e$ in which the possible sample sizes are 0 and 1, or 0 only. If 0 ane 1 are both possible sample sizes, two members of $\mathcal{C}'_e$ are A: take 0 observations and estimate 0 for θ; and B: take 1 observation and estimate x_1 for θ. Procedure A has risk function $r(\theta \mid A) = W(\theta)$. Procedure B has risk function $r(\theta \mid B) = c + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty W(y) e^{-\frac{1}{2}y^2} \, dy = \sup_y W(y)$. It easily follows that, except for A, all members of $\mathcal{C}'_e$ are non-admissible. The procedure A is admissible. For let S be any procedure which assigns probability $\alpha > 0$ to sample sizes > 0. Then we have

$$r(0 \mid S) \geq \alpha c + W(0) > W(0) = r(0 \mid A),$$

so that no such S could make A non-admissible. Let T be any procedure which assigns probability 1 to the sample size 0. For any such procedure the risk $r(\theta \mid T)$ is an average, for some distribution of z, of $W(z - \theta)$. Let $(-\theta_0, \theta_0)$ be the interval or point on which $W(\theta) = W(0)$. Clearly we cannot have $r(\theta \mid T) = W(0)$ for all $\theta \, \epsilon \, (-\theta_0, \theta_0)$ unless T coincides with A with probability 1. Hence no such T could make A non-admissible, and it now follows that A is admissible. This proof also shows that A is admissible when 0 is the only possible sample size for members of $\mathcal{C}'_e$.

Similar arguments show that every member of $\mathfrak{C}$ which belongs to a $\mathfrak{C}'_e$ of the above types, is admissible for the problems (i) and (ii) for which it is minimax.

6. Extensions of normal example. An outline of the solution of the example of section 5 shows that the same method can be used for other examples. Let X_1, X_2, $\cdots$ be independent random variables, each having the same density $p_\theta(x)$ with respect to a fixed measure μ. A point estimate z is wanted for the real parameter θ. Let

$$W_1(n, z, \theta) = W_1(n), \; W_2(n, z, \theta) = W_2(z, \theta)$$

and define the three classes of minimax problems as usual.

Suppose that we can find a sequence ξ_1, ξ_2, $\cdots$ of *a priori* distributions and a double sequence $z_{k,0}$, $z_{k,1}(x_1)$, $z_{k,2}(x_1, x_2)$, $\cdots$; $k = 1, 2, \cdots$ of estimates of θ, such that if θ has *a priori* distribution ξ_k and we observe $x_1, \cdots, x_m$ and then stop, the *a posteriori* risk is minimised by estimating $z_{k,m}$ for θ, and the minimum value is

$$r_{e,k}(m) = cW_1(m) + \int_{-\infty}^{\infty} W_2(z_{k,m}, \theta)p(\theta \mid x_1, \cdots, x_m ; \xi_k) \, d\theta,$$

depending not on the observations but only on the number m of observations (and c, k). Clearly the same sequences will do for all c, $0 < c < \infty$, and for all functions $W_1(n)$.

Then the following procedures are Bayes for the problem (iii)' with this c, and with θ having *a priori* distribution ξ_k: The only possible sample sizes are those which minimise $r_{e,k}(m)$: $m = 0, 1, 2, \cdots$; if the sample size is m estimate $z_{k,m}$ for θ.

Suppose for a particular ξ_k and for some particular c, that these possible sample sizes are $n_1 < n_2 < \cdots$. Since $r_{e,k}(m)$ is continuous in c for any k, m it is clear that for ε sufficiently small and $c < c' < c + \varepsilon$, no value of m other than n_1, n_2, $\cdots$ could minimise $r_{e',k}(m)$: $m = 0, 1, 2, \cdots$. And a minimum for any $m > n_1$ would provide a contradiction of Lemma 3. Hence for $c < c' < c + \varepsilon$, $r_{e',k}(m)$: $m = 0, 1, 2, \cdots$ is minimised by $m = n_1$ only. It follows that randomisations in sample size for Bayes solutions are possible only for a denumerable set of c's; for almost all c, only one fixed sample size is possible.

Suppose that as $k \to \infty$ every sequence $z_{1,m}$, $z_{2,m}$, $\cdots$ tends to a limit z_m, and that $r_{e,k}(m) \to r_e(m) = cW_1(m) + L(m)$, for $m = 0, 1, 2, \cdots$. If the procedure: take a sample of fixed size m and estimate z_m for θ has risk function $r_e^*(\theta)$ $= cW_1(m) + L_\theta(m) \leq r_e(m)$ for all θ, the following procedures are minimax for the problem (iii) with this c: The only possible sample sizes are those which minimise $r_e(m)$: $m = 0, 1, 2, \cdots$. If the sample size is m estimate z_m for θ.

Extension to problems (i) and (ii) can now be carried out as in section 5. We note that a problem of this type when solved for one $W_1(m)$ can be solved for any other $W_1(m)$ by merely determining the proper sample sizes. If $r_e(m)$ is a convex function of m, the possible sample sizes are always one integer or two

consecutive integers. But if $r_c(m)$ is not convex, practically any set of integers can be possible sizes, as indicated in the following examples.

EXAMPLE. Let X_1, X_2, $\cdots$ be independent random variables, each being $N(\theta, 1)$. A point estimate z is wanted for the mean θ. Let

$$W_1(n) = \frac{1}{3} n \qquad \text{for} \quad n = 0, 1, 2, 3,$$

$$= 1 + \frac{n - 3}{105} \quad \text{for} \quad n = 4, 5, \cdots,$$

$$W_2(z, \theta) = (z - \theta)^2.$$

Thus the first three observations each cost $\frac{1}{3}$, subsequent observations each cost $\frac{1}{105}$. Making the necessary substitutions in section 5, we get

$$r_c(m) = c\,\frac{m}{3} + \frac{1}{m} \qquad \text{for} \quad m = 1, 2, 3,$$

$$= c + \frac{c(m - 3)}{105} + \frac{1}{m} \quad \text{for} \quad m = 4, 5, \cdots.$$

For $c = 1$ it is easy to show that $r_1(m): m = 1, 2, \cdots$ is minimised by $m = 2$ and $m = 10$. For $c \neq 1$, $r_c(m): m = 1, 2, \cdots$ is minimised by one integer or by a pair of consecutive integers. Solutions are obtained for all problems (i), (ii), (iii) as in section 5. The solution obtained for any problem (i) with $\frac{2}{3} \leq M \leq \frac{18}{15}$ is the following:

with probability $\dfrac{16 - 15M}{6}$ take 2 observations,

with probability $\dfrac{15M - 10}{6}$ take 10 observations,

estimate $z = \dfrac{1}{n} \sum_{i=1}^{n} x_i$ for θ.

EXAMPLE. Let X_1, X_2, $\cdots$ be independent random variables each being $N(\theta, 1)$. A point estimate z is wanted for the mean θ. Let

$$W_1(n) = 1 - \frac{1}{n} \quad \text{for} \quad n = 1, 2, \cdots,$$

$$= 0 \qquad \text{for} \quad n = 0,$$

$$W_2(z, \theta) = (z - \theta)^2.$$

Making the necessary substitutions in Section 5,

$$r_c(m) = c + (1 - c)\,\frac{1}{m} \quad \text{for} \quad m = 1, 2, \cdots.$$

Clearly $r_1(m): m = 1, 2, \cdots$ is constant. Thus any procedure in which the sample size is at least 1 and the estimate $z = \dfrac{1}{n} \sum_{i=1}^{n} x_i$ is used for θ, is minimax for the problem (iii) with $c = 1$. If $c < 1$, problem (iii) has no solution. (The larger the sample size, the smaller is the risk.) If $c > 1$, $r_c(m): m = 1, 2, \cdots$ is minimised by $m = 1$ only. (In both these examples the possibility $n = 0$ is excluded because $\sup_{\theta}$ (risk) is then ∞.)

7. Example: rectangular. Let $X_1, X_2, \cdots$ be a sequence of independent random variables, each being $R(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, i.e., rectangular with range $\theta - \frac{1}{2}$ to $\theta + \frac{1}{2}$. A point estimate z is wanted for the parameter θ. Let

$$W_1(n, z, \theta) = n, \qquad W_2(n, z, \theta) = W(z - \theta),$$

where W is a non-decreasing function of $|z - \theta|$. The three classes of minimax problems are

(i)   subject to $\sup_{\theta} E_{\theta}(n) \leq M,$ minimise $\sup_{\theta} E_{\theta} W(z - \theta),$

(ii)   subject to $\sup_{\theta} E_{\theta} W(z - \theta) \leqq L,$ minimise $\sup_{\theta} E_{\theta}(n),$

(iii)   minimise $\sup_{\theta} \{cE_{\theta}(n) + E_{\theta} W(z - \theta)\}.$

NOTE. The problems (iii) are solved by Wald in [3] for the case $W(z - \theta) = (z - \theta)^2$. We will show that Wald's solution holds for any $W(z - \theta)$ which is non-decreasing in $|z - \theta|$, and will obtain solutions of (i) and (ii). In addition, admissibility results will be proved as in Section 5.

The remainder of this section will be concerned with proving the following results.

THEOREM. *The following procedures are admissible solutions of problem* (iii) *among the class of all procedures with continuous risk functions. If* $\phi^* = \sup_{\alpha}$

$W(\alpha) - 2 \displaystyle\int_0^{\frac{1}{2}} W(\alpha)\, d\alpha - c < 0$ *take* 0 *observations and estimate* 0 *for* θ. *If* $\phi^* > 0$

take at least one observation and after the mth *observation* $(m = 1, 2, \cdots)$ *compute the range* r_m *of* $x_1, \cdots, x_m$. *If* $r_m > 1 - \bar{l}$ *stop and estimate the mid-range for* θ; *if* $r_m < 1 - \bar{l}$ *take another observation; if* $r_m = 1 - \bar{l}$ *do either. If* $\phi^* = 0$ *follow either procedure.* (*Here* $\bar{l}$, *to be defined later, is a constant depending on* c *and* W.) *The class of procedures so obtained,* $0 < c < \infty$, *provides admissible solutions among the class of procedures with continuous risk functions, for all problems* (i) *and* (ii).

Solutions are found for problems (iii) by first finding Bayes solutions for the corresponding problems (iii)′ when θ has *a priori* distribution $R(a, b)$. The Bayes problem is to find a sequential estimation procedure which minimises the risk

$$E\{r_c(\theta \mid S) \mid \theta \sim R(a, b)\} = \frac{1}{b - a} \int_a^b \{cE_{\theta}(n \mid S) + E_{\theta}(W \mid S)\}\, d\theta.$$

Let us first determine what should be our estimate z if we stop after having observed $x_1, \cdots, x_m$. For this we will need to know the *a posteriori* distribution $p(\theta \mid x_1, \cdots, x_m)$. Writing $u_m = \min (x_1, \cdots, x_m)$ and $v_m = \max (x_1, \cdots, x_m)$, this distribution is easily found to be $R(u'_m, v'_m)$, where $(u'_m, v'_m) = (a, b) \cap (v_m - \frac{1}{2}, u_m + \frac{1}{2})$ for $m = 1, 2, \cdots$, and $(u'_0, v'_0) = (a, b)$. Clearly a best estimate, i.e., one minimising the *a posteriori* risk $cm + \int W(z - \theta)p(\theta \mid x_1, \cdots, x_m) \, d\theta$ is $z = \dfrac{u'_m + v'_m}{2}$, the mid-point of (u'_m, v'_m). The minimum value is

$$r_m = cm + \frac{1}{v'_m - u'_m} \int_{u'_m}^{v'_m} W\left(\theta - \frac{u'_m + v'_m}{2}\right) d\theta$$

$$= cm + \frac{2}{t_m} \int_0^{\frac{1}{2}t_m} W(\alpha) \, d\alpha,$$

where $t_m = v'_m - u'_m$ for $m = 0, 1, 2, \cdots$.

To determine an optimum stopping rule we will need to know, for all $t > 0$, the conditional expected value of r_{m+1} given $t_m = t$. Now

$$p(x_{m+1} \mid t_m = t) = \frac{1}{t}\left(\text{length of } \left(\frac{u'_m + v'_m}{2} - \frac{t}{2}, \frac{u'_m + v'_m}{2} + \frac{t}{2}\right) \right.$$

$$\left. \cap \ (x_{m+1} - \tfrac{1}{2}, x_{m+1} + \tfrac{1}{2})\right).$$

From this it is easy to show that

$$E(r_{m+1} \mid t_m = t) = c(m + 1) + \frac{2(t - 1)}{t} \int_0^{1/2} W(\alpha) \, d\alpha$$

$$+ \frac{4}{t} \int_0^1 \left[\int_0^{x/2} W(\alpha) \, d\alpha\right] dx$$

for $t \geqq 1$; and that for $t \leqq 1$,

$$E(r_{m+1} \mid t_m = t) = c(m + 1) + \frac{2(1 - t)}{t} \int_0^{t/2} W(\alpha) \, d\alpha$$

$$+ \frac{4}{t} \int_0^t \left[\int_0^{x/2} W(\alpha) \, d\alpha\right] dx.$$

Let

$$\phi(t) = cm + \frac{2}{t} \int_0^{t/2} W(\alpha) \, d\alpha - E(r_{m+1} \mid t_m = t),$$

the expected decrease in *a posteriori* risk due to taking $m + 1$ instead of m observations when $t_m = t$. We have

$$\theta(t) = \frac{2}{t} \int_0^{t/2} W(\alpha) \, d\alpha + \left(\frac{2}{t} - 2\right) \int_0^{1/2} W(\alpha) \, d\alpha - \frac{4}{t} \int_0^1 \left[\int_0^{x/2} W(\alpha) \, d\alpha\right] dx - c$$

for $t \geqq 1$; and for $t \leqq 1$,

$$\phi(t) = 2 \int_0^{t/2} W(\alpha) \, d\alpha - \frac{4}{t} \int_0^t \left[\int_0^{x/2} W(\alpha) \, d\alpha \right] dx - c.$$

Now $W(\alpha)$, being non-decreasing for $\alpha \geqq 0$, has at most a denumerable set of discontinuities. If $W(\alpha)$ is continuous at $\alpha = t/2$ we have, for $t > 1$:

$$\phi'(t) = \frac{1}{t} W\left(\frac{t}{2}\right) - \frac{2}{t^2} \int_0^{t/2} W(\alpha) \, d\alpha - \frac{2}{t^2} \int_0^{1/2} W(\alpha) \, d\alpha$$

$$+ \frac{4}{t^2} \int_0^1 \left[\int_0^{x/2} W(\alpha) \, d\alpha \right] dx$$

$$= \frac{1}{t} W\left(\frac{t}{2}\right) - \frac{2}{t^2} \int_0^1 \left[\int_0^{1/2} W(\alpha) \, d\alpha \right] dx - \frac{2}{t^2} \int_{1/2}^{t/2} W(\alpha) \, d\alpha$$

$$- \frac{2}{t^2} \int_0^1 \left[\int_0^{1/2} W(\alpha) \, d\alpha \right] dx + \frac{4}{t^2} \int_0^1 \left[\int_0^{x/2} W(\alpha) \, d\alpha \right] dx$$

$$= \frac{1}{t} W\left(\frac{t}{2}\right) - \frac{2}{t^2} \int_{1/2}^{t/2} W(\alpha) \, d\alpha - \frac{4}{t^2} \int_0^1 \left[\int_{x/2}^{1/2} W(\alpha) \, d\alpha \right] dx$$

$$\geqq \frac{1}{t} W\left(\frac{t}{2}\right) - \frac{2}{t^2} \frac{t-1}{2} W\left(\frac{t}{2}\right) - \frac{4}{t^2} W\left(\tfrac{1}{2}\right) \cdot \tfrac{1}{2}$$

$$= \frac{1}{t^2} W\left(\frac{t}{2}\right) - \frac{1}{t^2} W\left(\tfrac{1}{2}\right) \geqq 0;$$

and if $t < 1$ we have

$$\phi'(t) = W\left(\frac{t}{2}\right) - \frac{4}{t} \int_0^{t/2} W(\alpha) \, d\alpha + \frac{4}{t^2} \int_0^t \left[\int_0^{x/2} W(\alpha) \, d\alpha \right] dx$$

$$= W\left(\frac{t}{2}\right) - \frac{4}{t^2} \int_0^t \left[\int_0^{t/2} W(\alpha) \, d\alpha \right] dx + \frac{4}{t^2} \int_0^t \left[\int_0^{x/2} W(\alpha) \, d\alpha \right] dx$$

$$= W\left(\frac{t}{2}\right) - \frac{4}{t^2} \int_0^t \left[\int_{x/2}^{t/2} W(\alpha) \, d\alpha \right] dx$$

$$\geqq W\left(\frac{t}{2}\right) - \frac{4}{t^2} W\left(\frac{t}{2}\right) \frac{t^2}{4} = 0.$$

If $t/2$ is a discontinuity point of $W(\alpha)$, the same inequalities hold for the one-sided derivatives of $\phi(t)$, both of which exist. We observe that these inequalities are strict unless $W(\alpha)$ is constant on the open interval $(0, t/2)$. Noting that $\phi(t) \to -c$ as $t \to 0$, we have proved that $\phi(t)$ is continuous and non-decreasing for $t > 0$, being strictly increasing whenever $\phi(t) > -c$.

Hence $\phi(t) < 0$ for all t, or else $\phi(t) = 0$ has a unique root $\bar{t}$. Using also the fact that $t_{m+1} \leqq t_m$, we now obtain, by the methods of [3], the following results.

If $\phi(t) < 1$ for all t, a Bayes solution is: Take 0 observations, estimate $\dfrac{a+b}{2}$

for θ. If $\phi(t) > 0$ for some t, a Bayes solution is: After the mth observation $(m = 0, 1, 2 \cdots)$ compute $t_m = v'_m - u'_m$. If $t_m < \bar{l}$ stop and estimate $\dfrac{u'_m + v'_m}{2}$ for θ; if $t_m > \bar{l}$ take another observation; if $t_m = \bar{l}$ do either.

Consider now the following procedures S_0: If $\phi^* = \sup_a W(\alpha) - 2 \int_0^{1/2} W(\alpha) \, d\alpha - c < 0$ take 0 observations and estimate 0 for θ. If $\phi^* > 0$ take at least one observation, and after each observation $(m = 1, 2, \cdots)$ compute $t_m^* = u_m + \frac{1}{2} - (v_m - \frac{1}{2}) = u_m - v_m + 1$; if $t_m^* < \bar{l}$ stop and estimate $\dfrac{u_m + v_m}{2}$ for θ, if $t_m^* > \bar{l}$ take another observation, and if $t_m^* = \bar{l}$ do either. Finally, if $\phi^* = 0$ use either of these two procedures.

If $\phi^* > 0$ it is easy to show that $E_\theta(n \mid S_0)$, $E_\theta(W \mid S_0)$ and

$$r(\theta \mid S_0) = c E_\theta(n \mid S_0) + E_\theta(W \mid S_0) \equiv r$$

are all constants. Also, for any particular c, there is always an S_0 for which $E_\theta(n \mid S_0)$ is constant.

Let S_k be a Bayes procedure when θ has the distribution $\xi_k = R(-k, k)$ a priori. If $\phi^* \leq 0$, then for all k the procedure S_k is: take 0 observations and estimate 0 for θ; it thus coincides with an S_0. (Other possible S_0 have the same $\sup_\theta r(\theta \mid S_0)$.) If $\phi^* > 0$, then for all k sufficiently large the procedure S_k coincides with S_0 for $-(k - 1) \leq \theta \leq k - 1$. Taking a sequence of S_k's with $k \to \infty$, it easily follows from Lemma 6 that all procedures S_0 are minimax for the problem (iii) in question.

By the same methods as are used in section 5 it is easy to show that the procedures S_0 obtained above provide solutions for all solvable problems (i) and (ii).

In the case $\phi^* > 0$, for the procedure S_0 to be non-admissible for the problem (iii) for which it is minimax, there must exist a procedure S_0^* having risk function

$$r(\theta \mid S_0^*) \leq r \quad \text{for all} \quad \theta$$

with strict inequality for at least one θ and so, if $r(\theta \mid S_0^*)$ is continuous, for an interval of values of θ. We can therefore find $h > \frac{1}{2}$ such that

$$\frac{1}{2h - 1} \int_{-h+1/2}^{h-1/2} r(\theta \mid S_0^*) \, d\theta = a < r.$$

Now for $\alpha = \pm 2, \pm 4, \cdots$ define the procedure S_α^* as follows. If $x_1, x_2, \cdots$ are observed, use the decision procedure S_0^* for the sequence $x_1 - \alpha h$, $x_2 - \alpha h, \cdots$ and add αh to the resulting estimate. We clearly have

$$r(\theta \mid S_\alpha^*) = r(\theta - \alpha h \mid S_0^*).$$

Now define the procedure S^* as follows. Take at least one observation. If $x_1 \epsilon \overline{(\alpha - 1h}, \overline{\alpha + 1h}]$, $\alpha = 0, \pm 2, \pm 4, \cdots$, use the procedure S_α^*. If $\theta \epsilon \overline{(\alpha - 1h} + \frac{1}{2}, \overline{\alpha + 1h} - \frac{1}{2})$, then $x_1 \epsilon \overline{(\alpha - 1h}, \overline{\alpha + 1h}]$ and so the procedure S^* reduces to S_α^*. Hence $r(\theta \mid S^*)$ coincides with $r(\theta \mid S_\alpha^*)$ for

$$\theta \epsilon \overline{(\alpha - 1h} + \tfrac{1}{2}, \overline{\alpha + 1h} - \tfrac{1}{2}), \qquad \alpha = 0, \pm 2, \pm 4, \cdots.$$

And $r(\theta \mid S^*) \leqq r$ for all θ. Therefore

$$\frac{1}{2(2k + 1)h} \int_{-(2k+1)h}^{(2k+1)h} r(\theta \mid S^*) \, d\theta \leqq \frac{(2h - 1)a + r}{2h} = r - \frac{(r - a)(2h - 1)}{2h}.$$

But if θ has the distribution $R(-\overline{2k + 1}h, \overline{2k + 1}h)$ *a priori*, the Bayes solution coincides with S_0 for $\theta \epsilon (-\overline{2k + 1}h + 1, \overline{2k + 1}h - 1)$. We therefore have for this *a priori* distribution

$$\text{Bayes risk} \geqq \frac{2(2k + 1)h - 2}{2(2k + 1)h} r = r - \frac{r}{(2k + 1)h}.$$

For k sufficiently large this clearly exceeds $r - (2h - 1)(r - a)/2h$, contradicting the above inequality on the Bayes risk. It follows that no such S_0^* as assumed can exist and therefore that the procedure S_0 is admissible, among the class of procedures with continuous risk functions, for the problems (iii) for which it is minimax and also, by Lemma 7, for the problems (i) and (ii) for which it is minimax.

If W is continuous and the number of observations is bounded it can be shown that $r^*(\theta)$ is continuous. Thus if W is continuous, S_0 is admissible, among the class of procedures with n bounded, for the three problems.

It remains to consider admissibility for procedures S_0 where $\phi^* \leqq 0$. Proofs for these cases can be given in the same way as for the corresponding cases in Section 5.

The solution for this example still works if we replace $W_1(n) = n$ by some other $W_1(n)$, but only so long as the resulting function $\phi(t)$ is non-decreasing.

NOTE. In the above examples, a procedure is called cogredient if for every c the same number of observations is taken for $x_1 + c, x_2 + c, \cdots$ as for $x_1, x_2, \cdots$ and $z(x_1 + c, \cdots, x_n + c) = z(x_1, \cdots, x_n) + c$. Such procedures have constant risk functions; so it follows that all the constant risk procedures obtained in Sections 5, 6, 7 have uniformly minimum risk among all cogredient procedures for the problems for which they are minimax.

The author wishes to express his appreciation to Professor E. L. Lehmann for valuable advice and assistance in the preparation of this paper, and also to Jack C. Kiefer, who read the manuscript and pointed out an important correction.

REFERENCES

[1] C. STEIN AND A. WALD, "Sequential confidence intervals for the mean of a normal
         distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp.
         427–433.
[2] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known
         variance," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 218–230.
[3] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[4] J. HODGES AND E. LEHMANN, "Extensions and applications of the Cramér-Rao in-
         equality," *Proceedings of the Second Berkeley Symposium on Mathematical Sta-
         tistics and Probability*, University of California Press, 1951.

ON MINIMUM VARIANCE IN NONREGULAR ESTIMATION

By R. C. Davis

U. S. Naval Ordnance Test Station, China Lake, California

1. Summary. A case of nonregular estimation arises in attempting to estimate a single unknown parameter, θ, in the probability distribution of a single chance variable in which one or both of the extremities of the range of the distribution are functions of the unknown parameter. The case treated in this paper is the one in which a probability density of exponential type exists. When one extremity alone of the range depends non-trivially upon θ, a necessary and sufficient condition is given in order that a single order statistic be a sufficient statistic for θ. In this case conditions are given for the existence of a unique unbiased estimate of θ possessing minimum variance uniformly in θ. In the case in which both extremities of the range depend upon θ, a necessary and sufficient condition is given that the smallest and largest order statistics constitute a set of sufficient statistics for θ. In this case Pitman [1] has shown that a single sufficient statistic exists if one extremity of the range is a monotone decreasing function of the other extremity.[1] It is shown that under the above condition a unique unbiased estimate exists possessing minimum variance. Moreover a surmise of Pitman is proved that only under this condition does a single sufficient statistic exist. When a single sufficient statistic does not exist, an unbiased estimate of a known function of θ is obtained which has less variance than any analytic function of the set of sufficient statistics for θ.

2. Introduction. Let X be a chance variable assuming values x in a one-dimensional Euclidean space, R_1, and let X possess a probability density function $f(x, \theta)$ depending on a single unknown parameter θ which lies in Ω, a subset of R_1. Denote by $a(\theta)$ and $b(\theta)$ the lower and upper extremities of the range of $f(x, \theta)$. We treat the cases in which either one or both the extremities of the range depend nontrivially upon θ. For each $\theta \, \epsilon \, \Omega$ denote by $R^*(\theta)$ the subset of R_1 satisfying $a(\theta) \leqq x \leqq b(\theta)$, and by $R^{**}(\theta)$ the complement of $R^*(\theta)$ in R_1. We make the following assumptions:

ASSUMPTION A.

$$f(x, \theta) = 0 \quad \text{for all} \quad (x, \theta) \text{ on } R^{**}(\theta) \times \Omega;$$

$$f(x, \theta) = e^{\theta K(x) + S(x) + T(\theta)} \quad \text{for all} \quad (x, \theta) \text{ on } R^*(\theta) \times \Omega,$$

where $T(\theta)$ is a real single-valued continuous function of θ at all points of Ω, and $S(x)$, $K(x)$ are real single-valued continuous functions of x defined almost everywhere on R_1.

ASSUMPTION B. $a(\theta)$ and $b(\theta)$ are continuous functions of θ satisfying for all $\theta \, \epsilon \, \Omega$ the inequality $a(\theta) \leqq b(\theta)$.

---

[1] The author is deeply indebted to the referee for bringing to his attention the paper by Pitman and for many other helpful suggestions.

The exponential type of frequency function assumed above is the type which Koopman [2] has shown to hold whenever a sufficient statistic for θ exists. We do not require any of his results, however, in this paper.

For convenience in notation we write $P(x) = e^{S(x)}$ and $Q(\theta) = e^{T(\theta)}$, so that obviously we have the relation

$$[Q(\theta)]^{-1} = \int_{a(\theta)}^{b(\theta)} P(\eta) \, d\eta.$$

Furthermore if an estimate of θ is a continuous function of n independent sample values, is unbiased, and possesses minimum variance uniformly in $\theta \, \epsilon \, \Omega$, we term this a best estimate of θ.

3. One extremity of the range depending upon θ. First we treat the case in which only one extremity of the range depends upon the unknown parameter θ. To fix the argument we assume that the upper extremity $b(\theta)$ depends upon θ, and the lower extremity is independent of θ. The results of this section are extended in an obvious manner to the case in which the lower extremity alone depends upon θ.

THEOREM 1. *Let x_1, x_2, $\cdots$, x_n be the values of n independent drawings from a population having the probability density function $f(x, \theta)$ satisfying Assumptions A and B, and in which the upper extremity only of the range depends upon θ. The necessary and sufficient condition that the nth order statistic, denoted by $x_{(n)}$, be a sufficient statistic for θ is that*

$$f(x, \theta) = P(x) \, Q(\theta) \quad \text{for all} \quad (x, \theta) \text{ in } R^*(\theta) \times \Omega.$$

PROOF OF NECESSITY. Suppose that in a sample of n independent observations that the nth order statistic, $x_{(n)}$, is a sufficient statistic for θ. It follows from the definition of sufficiency that

$$f(x_1, \theta) \, \cdots \, f(x_n, \theta) = g(x_{(n)}; \theta) \, h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(n)}; \theta),$$

where $g(x_{(n)}, \theta)$ is the frequency function of $x_{(n)}$, and $h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(n)}; \theta)$ denotes the conditional frequency function of the order statistics $x_{(1)}, \cdots, x_{(n-1)}$, given a fixed value of $x_{(n)}$, and is independent of θ. It is well known from the theory of order statistics that $g(x_{(n)}; \theta)$ has the form

$$g(x_{(n)}; \theta) = n[F(x_{(n)})]^{n-1} f(x_{(n)}) = nP(x_{(n)})[Q(\theta)]^n e^{\theta K(x_{(n)})} \left[\int_a^{x_{(n)}} P(\eta) e^{\theta K(\eta)} \, d\eta \right]^{n-1},$$

where $F(x_{(n)}) = \int_a^{x_{(n)}} f(\eta, \theta) \, d\eta.$

It follows from the above that

$$(1) \qquad h(x_{(1)}, \cdots, x_{(n-1)} \mid x_n; \theta) = \frac{\left[\exp \left[\theta \sum_{i=1}^{n-1} K(x_{(i)}) \right] \right] \prod_{j=1}^{n-1} P(x_j)}{n \left[\int_a^{x_{(n)}} P(\eta) e^{\theta K(\eta)} \, d\eta \right]^{n-1}},$$

where $h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(n)}; \theta)$ is independent of θ. Differentiating equation (1) partially with respect to θ, substituting the value of $h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(n)}; \theta)$ from (1) and placing $\frac{\partial h}{\partial \theta} = 0$, we obtain after some simple algebra

$$(2) \qquad \int_a^{x_{(n)}} K(\eta) P(\eta) e^{\theta K(\eta)} \, d\eta = \frac{\left[\sum_{i=1}^{n-1} K(x_{(i)}) \right]}{n-1} \int_a^{x_{(n)}} P(\eta) e^{\theta K(\eta)} \, d\eta.$$

Since $f(x, \theta) \geq 0$ for all x in R_1, it follows that $P(\eta) e^{\theta K(\eta)} \geq 0$, for $a \leq \eta \leq x_{(n)}$. Moreover we obtain from the first mean value theorem for integrals that

$$\int_a^{x_{(n)}} K(\eta) P(\eta) e^{\theta K(\eta)} \, d\eta = K(\xi) \int_a^{x_{(n)}} P(\eta) e^{\theta K(\eta)} \, d\eta,$$

where $a \leq \xi \leq x_{(n)}$. Equation (2) reduces then to the form

$$(3) \qquad K(\xi) = \frac{1}{n-1} \sum_{i=1}^{n-1} K(x_{(i)}).$$

It is noted that the only sample value on which ξ is dependent is the $x_{(n)}$. Equation (3) is valid for every $x_{(1)}, \cdots, x_{(n-1)}$, satisfying the inequalities $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}$ with the $x_{(i)}$ assuming values in $R^*(\theta)$. Let $x_{(n)}$ take some fixed value arbitrarily close to $b(\theta)$. (If $f(b(\theta), \theta) \neq 0$, we can of course let $x_{(n)} = b(\theta)$.) Also let x be any number satisfying the inequality $a \leq x \leq x_n$. Now let $x_{(1)} = x_{(2)} = \cdots = x_{(n-1)} = x$, and we obtain from (3) the relation (4) $K(x) = K(\xi)$.

Since this relation is true for every x in the interval $a \leq x < x_{(n)}$, it follows that $K(x)$ is a constant in the interval $a \leq x < x_{(n)}$. (Again if we assume $f(b(\theta), \theta) \neq 0$, we can let $x_{(n)} = b(\theta)$, and it follows that $K(x)$ would be a constant in the closed interval $a \leq x \leq b(\theta)$.) Therefore, necessity is proved.

PROOF OF SUFFICIENCY. This proof is extremely simple. If $f(x, \theta) = P(x) Q(\theta)$, we have

$$h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(n)}; \theta) = \frac{[Q(\theta)]^n \, P(x_{(1)}) \cdots P(x_{(n)})}{n[Q(\theta)]^n P(x_n) \left[\int_a^{x_n} P(\eta) \, d\eta \right]^{n-1}}$$

and is independent of θ. Hence $x_{(n)}$ is a sufficient statistic for θ. This completes the proof of Theorem 1.

Before proceeding to the problem of constructing a best estimate for θ, we will use a theorem due to Blackwell [3] which will enable us to restrict ourselves to the class of unbiased estimates of θ which are functions of the sufficient statistic for θ. Blackwell's results are applicable to a much more general situation than we are considering here, and the results needed can be obtained in a different manner. Nevertheless we will summarize briefly the result which we need. He has proved that if x is any chance variable and y is any numerical

chance variable for which $E(y)$ and $E[y - E(y)]^2$ are finite, and $f(x)$ is any real valued function for which $E[f(x)y]$ is finite, then $\sigma^2 E(y \mid x)$ is finite, where $E(y \mid x)$ denotes the conditional expected value of y given x. Moreover he proves that $E[f(x) \, E(y \mid x)] = E[f(x)y]$ and $\sigma^2 E(y \mid x) \leqq \sigma^2 y$ with equality holding only if $y = E(y \mid x)$ with probability one.

As a particular application of Blackwell's result it follows that if a sufficient statistic S exists, and if t is any unbiased estimate of θ, then $\alpha(S) = E[t \mid S]$ is an unbiased estimate of θ with $\sigma^2[\alpha(S)] \leqq \sigma^2 t$. It follows that we can restrict ourselves (in the case in which only the upper extremity of the range depends on θ) to the class of functions of the sufficient statistic $x_{(n)}$ which yield sufficient statistics. If we can obtain out of this class a unique function of $x_{(n)}$ which is unbiased and possesses minimum variance in this class, we will obtain an unbiased estimate of θ possessing minimum variance.

4. Derivation of the best estimate for θ when the range varies from a to $b(\theta)$. If we make the transformation of parameters $\varphi = [Q(\theta)]^{-1}$, matters are simplified considerably. If we assume that the function $\varphi(\theta)$ possesses a unique inverse $\theta(\varphi)$ and let $c(\varphi) = b[\theta(\varphi)]$, we have the condition that $\alpha(x_{(n)})$ is an unbiased estimate of φ in the form

$$(5) \qquad \int_a^{c(\varphi)} \alpha(x_{(n)}) g(x_{(n)}, \varphi) \, dx_{(n)} \equiv \varphi.$$

This reduces to the condition

$$\int_a^{c(\varphi)} \alpha(x_{(n)}) P(x_{(n)}) \left[\int_a^{x_{(n)}} P(\eta) \, d\eta \right]^{n-1} dx_{(n)} = \frac{\varphi^{n+1}}{n}.$$

If we use a new variable of integration u, where $u = \int_a^{x_{(n)}} P(\eta) \, d\eta$, and let $\alpha(x_{(n)}) = \psi(u)$, the condition of unbiasedness becomes

$$\int_0^\varphi \psi(u) u^{n-1} \, du = \frac{\varphi^{n+1}}{n}.$$

Clearly the only solution of this integral equation which is an analytic function of u is given by

$$\psi(u) = \left(1 + \frac{1}{n} \right) u.$$

Since this is the unique solution for all finite φ, it follows that

$$\left(1 + \frac{1}{n} \right) \int_a^{x_{(n)}} P(\eta) \, d\eta$$

is the only unbiased estimate of φ. Its variance can be obtained by a simple integration, and we obtain

$$\sigma_\alpha^2 = \frac{\varphi^2}{n(n + 2)}.$$

If we wish to obtain an estimate for θ directly, the analysis is somewhat more complicated. Moreover it is necessary to make a further assumption to insure that the unique unbiased estimate of θ among the class of functions of $x_{(n)}$ is also a sufficient statistic. We may state this assumption as follows:

ASSUMPTION C. $b(\theta)$ *is a strictly monotone function of* θ. *If we define the following well defined functions*

$$u(x_{(n)}) = \int_a^{x_{(n)}} P(\eta) \, d\eta, \qquad \beta(x_{(n)}) = b^{-1}(x_{(n)}),$$

the functions $u(x_{(n)})$ *and* $\beta(x_{(n)})$ *satisfy the following condition:*

$$u \frac{d}{du} \left[\ln \left(\frac{d\beta}{du} \right) \right] > -2 \qquad (if \; b(\theta) \; is \; strictly \; monotone \; increasing),$$

$$u \frac{d}{du} \left[\ln \left(\frac{d\beta}{du} \right) \right] < -2 \qquad (if \; b(\theta) \; is \; strictly \; monotone \; decreasing).$$

Moreover, the parameter set Ω *is the interval defined by* $\theta \geq \theta_0$ *when* $b(\theta)$ *is strictly monotone increasing and the interval* $\theta \leq \theta_0$ *when* $b(\theta)$ *is strictly monotone decreasing.* θ_0 *satisfies the equation* $b(\theta) = \theta$, *so that* $b(\theta_0) = a$.

Let $\alpha(x_{(n)})$ represent now a function of the sufficient statistic $x_{(n)}$. The condition that α be an unbiased estimate is expressed in the form

$$(6) \qquad \int_a^{b(\theta)} \alpha(x_{(n)}) g(x_{(n)}, \theta) \, dx_{(n)} \equiv \theta$$

for every $\theta \, \epsilon \, \Omega$. This reduces to the condition

$$\int_a^{b(\theta)} \alpha(x_{(n)}) P(x_{(n)}) \left[\int_a^{x_{(n)}} P(\eta) \, d\eta \right]^{n-1} dx_{(n)} \equiv \frac{\theta}{n[Q(\theta)]^n}.$$

If we make the same substitution used before; namely, $u = \int_a^{x_{(n)}} P(\eta) \, d\eta$, and let $\alpha(x_{(n)}) = \psi(u)$, the condition of unbiasedness becomes

$$(7) \qquad \int_0^{1/Q(\theta)} \psi(u) u^{n-1} \, du \equiv \frac{\theta}{n[Q(\theta)]^n}.$$

It follows from Assumptions B and C that $\dfrac{db}{d\theta}$ and hence $\dfrac{dQ}{d\theta}$ exist almost everywhere in Ω. Hence differentiating (7), we obtain after simple algebra,

$$\psi \left[\frac{1}{Q(\theta)} \right] = \theta - \frac{1}{n \dfrac{d}{d\theta} \ln Q(\theta)}$$

for $\theta \, \epsilon \, \Omega$. Since Ω is an interval having θ_0 as one end point, we obtain after some manipulation the expression

$$(8) \qquad \alpha(x_{(n)}) = \beta(x_{(n)}) + \frac{u}{n} \frac{d\beta(x_{(n)})}{du(x_{(n)})},$$

where $\beta(x_{(n)})$ is the function inverse to $b(x_{(n)})$, denoted in Assumption C as $b^{-1}(x_{(n)})$. $\alpha(x_{(n)})$ is the only continuous function of $x_{(n)}$ which is an unbiased estimate of θ. In order to insure that $\alpha(x_{(n)})$ is also a sufficient statistic we must be certain that $\alpha(x_{(n)})$ has a unique inverse $\alpha^{-1}(x_{(n)})$. If we take the case in which $b(\theta)$ is strictly monotone increasing, this condition becomes

$$(9) \qquad \frac{d\alpha}{dx_{(n)}} = (n + 1)\frac{d\beta}{dx_{(n)}} + u\,\frac{du}{dx_{(n)}}\,\frac{d^2\beta}{d^2 u} > 0.$$

If Assumption C holds, $\alpha(x_{(n)})$ is a sufficient statistic for $n \geqq 1$. Finally applying Blackwell's theorem we conclude that $\alpha(x_{(n)})$ given by (8) is the best estimate of θ. From (9) it is obvious that if the function $u\,\dfrac{d}{du}\left[\ln\left(\dfrac{d\beta}{du}\right)\right]$ is a bounded function of $x_{(n)}$ for $a \leqq x_{(n)} \leqq b(\theta)$, $\theta \,\epsilon\, \Omega$, then for n sufficiently large $\alpha(x_{(n)})$ is a sufficient statistic and hence is the best estimate of θ assuming only the strict monotonicity of the function $b(\theta)$.

4a. *Examples.*

Rectangular Distribution. Let

$$f(x, \theta) = \frac{1}{\theta}, \qquad 0 \leqq x \leqq \theta,$$
$$= 0, \qquad \text{otherwise.}$$

Since $P(x) = 1$, and $b(\theta) = \theta$, we obtain $u = x_{(n)}$ and $\beta = x_{(n)}$. Hence

$$\alpha(x_{(n)}) = \beta(x_{(n)}) + \frac{u}{n}\frac{d\beta}{du} = \left(1 + \frac{1}{n}\right) x_{(n)}.$$

Its variance is given by the expression $\sigma_a^2 = \dfrac{\theta^2}{n(n + 2)}$.

Exponential Distribution. Let

$$f(x, \theta) = e^{x-\theta}, \qquad -\infty \leqq x \leqq \theta,$$
$$= 0, \qquad x > \theta.$$

Since $P(x) = e^x$, and $b(\theta) = \theta$, we obtain $u = e^{x_{(n)}}$, $\beta = x_{(n)}$. Hence

$$\alpha(x_{(n)}) = \beta(x_{(n)}) + \frac{u}{n}\frac{d\beta}{du} = x_{(n)} + \frac{1}{n}.$$

5. Both extremities of the range depending upon θ.

THEOREM 2. *Let $x_1, x_2, \cdots, x_n$ be the values of n independent drawings from a population having the probability density function $f(x, \theta)$ satisfying Assumptions A and B, and in which both extremities of the range depend upon θ. The necessary and sufficient condition that the first and nth order statistics, $x_{(1)}$ and $x_{(n)}$, be jointly sufficient statistics for θ is that*

$$f(x, \theta) = P(x)\,Q(\theta) \quad \text{for all} \quad (x, \theta) \text{ in } R^*(\theta) \times \Omega.$$

PROOF OF NECESSITY. Suppose that in a sample of n independent observations that the first and nth order statistics, $x_{(1)}$ and $x_{(n)}$, are jointly sufficient for θ. It follows from the definition of joint sufficiency that

$$f(x_1, \theta) \cdots f(x_n, \theta) = g(x_{(1)}, x_{(n)}; \theta)h(x_{(2)}, \cdots, x_{(n-1)} \mid x_{(1)}, x_{(n)}; \theta),$$

where $g(x_{(1)}, x_{(n)}; \theta)$ is the joint frequency function of $x_{(1)}$ and $x_{(n)}$, and

$$h(x_{(1)}, \cdots, x_{(n-1)} \mid x_{(1)}, x_{(n)}; \theta)$$

denotes the conditional frequency function of the order statistics $x_{(2)}, \cdots, x_{(n-1)}$, given fixed values of $x_{(1)}$ and $x_{(n)}$, and is independent of θ. It is well known from the theory of order statistics that $g(x_{(1)}, x_{(n)}; \theta)$ has the form

$$g(x_{(1)}, x_{(n)}; \theta) = n(n-1)[F(x_{(n)}) - F(x_{(1)})]^{n-2} f(x_{(1)}) f(x_{(n)}),$$

where $F(x_{(n)}) - F(x_{(1)}) = \int_{x_{(1)}}^{x_{(n)}} f(\eta, \theta) \, d\eta$. It follows from the above that

$$h(x_{(2)}, \cdots, x_{(n-1)}, \mid x_{(1)}, x_{(n)}; \theta) = \frac{\left[\exp\left[\theta \sum_{i=2}^{n-1} K(x_{(i)}) \right] \right] \prod_{j=2}^{n-1} P(x_{(j)})}{n(n-1)\left[\int_{(x_{(1)})}^{x_{(n)}} P(\eta) e^{\theta K(\eta)} \, d\eta \right]^{n-2}}.$$

The proof proceeds similarly to the one in Theorem 1, and we end up with a similar equation

$$(10) \qquad K(\xi) = \frac{1}{n-2} \sum_{i=2}^{n-1} K(x_{(i)}),$$

where $x_{(1)} \leqq \xi \leqq x_{(n)}$. Hence by a similar argument $K(x)$ is a constant in the open interval $a(\theta) < x < b(\theta)$. If $f(a(\theta), \theta)$ and $f(b(\theta), \theta)$ are unequal to zero, we can make the stronger statement that $K(x)$ is a constant in the closed interval $a(\theta) \leqq x \leqq b(\theta)$.

PROOF OF SUFFICIENCY. Suppose that $f(x, \theta) = P(x) Q(\theta)$. Then

$$h(x_{(2)}, \cdots, x_{(n-1)}, \mid x_{(1)}, x_{(n)}; \theta) = \frac{[Q(\theta)]^{n-2} \prod_{i=2}^{n-1} P(x_{(i)})}{n(n-1)[Q(\theta)]^{n-2}\left[\int_{x_{(1)}}^{x_{(n)}} P(\eta) \, d\eta \right]^{n-2}}$$

and is independent of θ. Hence $x_{(1)}$ and $x_{(n)}$ are jointly sufficient statistics for θ. This completes the proof of Theorem 2.

Blackwell's theorem is applicable again to this case and enables us to restrict ourselves to the class of unbiased estimates which are sufficient statistics for θ. Any unbiased sufficient statistic is a solution of the integral equation

$$(11) \qquad \int_{a(\theta)}^{b(\theta)} dx_{(n)} \int_{a(\theta)}^{x_{(n)}} \alpha(x_{(1)}, x_{(n)}) \, g(x_{(1)}, x_{(n)}) \, dx_{(1)} \, dx_{(n)} \equiv \theta$$

for $\theta \, \epsilon \, \Omega$.

Pitman has shown [1] that in the particular case $a(\theta) = \theta$, $b(\theta)$ a strictly monotone decreasing function of θ, a sufficient statistic for θ exists. An independent proof is given of this statement. Moreover, the distribution of this sufficient statistic is derived, and it is shown that there exists a unique unbiased estimate of θ in the class of all functions of the sufficient statistic.

Following Pitman we simplify the discussion considerably by assuming $a(\theta) = \theta$. On the basis of Theorem 2 and Blackwell's result we need only consider functions of the smallest and largest order statistics in our search for a best estimate. First we derive Pitman's result independently. Let us consider the sample statistic

$$T = \min. \{x_{(1)}, b^{-1}(x_{(n)})\}.$$

We proceed first to find its probability distribution and then show that it is a sufficient statistic for θ. Figure 1 shows a typical contour of constant T in the $x_{(1)}$, $x_{(n)}$ plane.
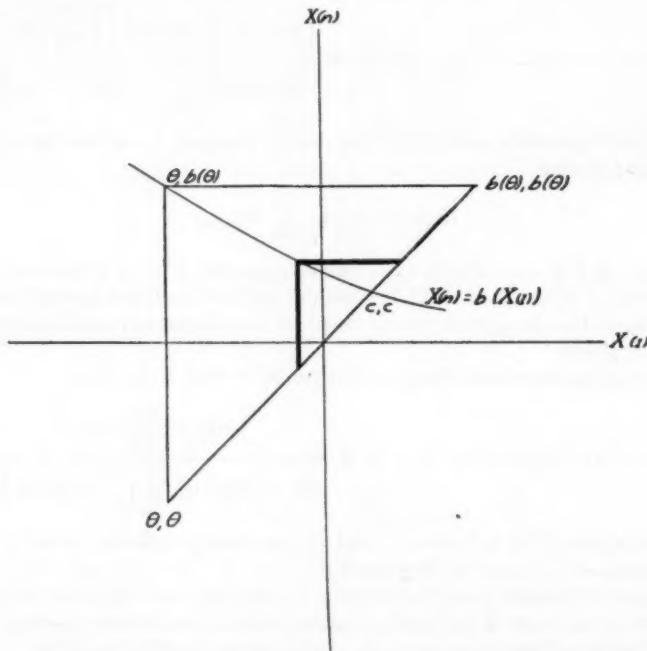
FIGURE 1

First it is clear from Assumption A that we may confine ourselves to the interior of the triangle shown in Fig. 1. Moreover, it is clear from the continu-

ity and monotony of the function $b(\theta)$ that there exists a point with coordinates c, c (where $b(c) = c$) which is independent of θ and is such that

$$\theta \leqq c \leqq b(\theta) \text{ for all } \theta \, \epsilon \, \Omega.$$

From Assumption B, $\Omega \subseteqq I$, where I is the interval in R_1 given by $\theta \leqq c$. It it clear from the definition of T that

$T \equiv b^{-1}(x_{(n)})$ for all points above the curve $x_{(n)} = b(x_{(1)})$,
$T \equiv x_{(1)}$ for all points below the curve $x_{(n)} = b(x_{(1)})$,
$T \equiv x_{(1)} \equiv b^{-1}(x_{(n)})$ for all points on the curve $x_{(n)} = b(x_{(1)})$.

A typical contour of constant T is shown in the figure. If we denote as before by $g(x_{(1)}, x_{(n)})$ the joint frequency function of the order statistics $x_{(1)}$ and $x_{(n)}$, it follows that

$$(12) \qquad Pr.\{t < T < t + dt\} = \left[\int_{t}^{b(t)} g(x_{(1)}, x_{(n)}) \, dx_{(n)} \right]_{[x_{(1)}=t]} dt$$

$$+ \left[\int_{t}^{b(t)} g(x_{(1)}, x_{(n)}) \, dx_{(1)} \right]_{[x_{(n)}=b(t)]} [b(t) - b(t + dt)],$$

where the first integral is evaluated holding $x_{(1)} = t$ and the second integral holding $x_{(n)} = b(t)$. It follows from the continuity and monotony of $b(\theta)$ that if we restrict the parameter set Ω to be a bounded interval in R_1, $\frac{db}{d\theta}$ will exist everywhere except on a set of points having probability measure zero. In this case T possesses a frequency function $w(t)$ almost everywhere. After performing the elementary integrations in (12) by noting that the integrands can be expressed as perfect differentials, we obtain

$$(13) \qquad w(t) = n[Q(\theta)]^n \left[\int_{t}^{b(t)} P(\eta) \, d\eta \right]^{n-1} \left[P(t) - \frac{db}{dt} P(b(t)) \right].$$

To prove that T is a sufficient statistic for θ, we must prove that the conditional frequency function of $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$, given T, is independent of θ. To do this we show that this property holds in each of the two regions indicated in Figure 1; namely in the regions below and above the curve $x_{(n)} = b(x_{(1)})$. In the region below the curve, we have

$$h(x_{(1)}, x_{(2)}, \cdots, x_{(n)} \mid T) = \frac{P(x_{(1)})P(x_{(2)}) \cdots P(x_{(n)})[Q(\theta)]^n}{w(t)}.$$

Obviously this conditional frequency function is independent of θ. In the region above the curve, $x_{(n)} = b(x_{(1)})$, we make the following transformation in the sample space: Let $\rho_1 = x_{(1)}, \rho_2 = x_{(2)}, \cdots, \rho_{n-1} = x_{(n-1)}, \rho_n = T$. Since

$$\frac{\partial(\rho_1, \rho_2, \cdots, \rho_n)}{\partial(x_{(1)}, x_{(2)}, \cdots, x_{(n)})} = \frac{db^{-1}(x_{(n)})}{dx_{(n)}},$$

the transformed likelihood function becomes

$$f(x_{(1)}, \theta)f(x_{(2)}, \theta) \cdots f(x_{(n)}, \theta) \cdot \left(\frac{db^{-1}}{dx_{(n)}}\right)^{-1}.$$

If we now assume that $b^{-1}(x_{(n)})$ is a strictly monotone decreasing function of $x_{(n)}$, the transformation is one-to-one and $\dfrac{db^{-1}}{dx_{(n)}}$ is unequal to zero except possibly at a set of points in the $x_{(1)}$, $x_{(n)}$ plane of probability measure zero. We may state then that

$$h(x_{(1)}, x_{(2)}, \cdots, x_{(n)} \mid T) = \frac{P(x_{(1)})P(x_{(2)}) \cdots P(x_{(n)})[Q(\theta)]^n \left(\dfrac{db^{-1}}{dx_{(n)}}\right)^{-1}}{w(t)}.$$

Again this conditional frequency function is independent of θ, so that this property holds throughout the triangle in Figure 1. Hence T is a sufficient statistic for θ.

We proceed to prove that there exists a unique continuous function of T which is an unbiased estimate of θ. This will involve no additional assumptions not made already. If $\psi(t)$ is an unbiased estimate of θ, we have from (13)

$$(14) \quad E[\psi(t)] = n[Q(\theta)]^n \int_\theta^c \psi(t)\left[\int_t^{b(t)} P(\eta)\, d\eta\right]^{n-1}\left[P(t) - \frac{db}{dt} P(b(t))\right] dt \equiv \theta$$

for $\theta \in \Omega$. Differentiating (14) with respect to θ, we obtain

$$\psi(\theta) = \theta - \frac{1}{n\dfrac{d}{d\theta} [\ln Q(\theta)]}.$$

Since Ω is the interval $\theta \leq c$, we obtain

$$(15) \qquad \psi(T) = T - \frac{1}{n\dfrac{d}{dt} [\ln Q(T)]}.$$

Hence (15) with $T = \min. \{x_{(1)}, b^{-1}(x_{(n)})\}$ is the unique continuous function of T which is an unbiased estimate of θ.

We now require an additional assumption to insure that $\psi(T)$ given by (15) is a sufficient statistic for θ.

ASSUMPTION D. *For almost all T satisfying $\theta \leq T \leq c$, and for all $\theta \in \Omega$, the function $\ln Q(T)$, where $[Q(T)]^{-1} = \int_T^{b(T)} P(\eta)\, d\eta$, satisfies the inequality*

$$-1 < \frac{\dfrac{d^2}{dt^2} [\ln Q(T)]}{\left[\dfrac{d}{dt} \ln Q(T)\right]^2} < M,$$

where M is some fixed constant.

The following theorem can be established:

THEOREM 3. *If a probability distribution with range from θ to $b(\theta)$ satisfies Assumptions A, B, and D, with $K(x) \equiv 0$, and if the functions $b(\theta)$ and $b^{-1}(\theta)$ are strictly monotone decreasing for all $\theta \, \epsilon \, \Omega$, then the function $\psi(T)$ given by (15), where $T = min. \{x_{(1)}, b^{-1}(x_{(n)})\}$, is the unique best estimate for the unknown parameter θ.*

PROOF. Under the above assumptions (minus Assumption D) we have proved that $\psi(T)$ given by (15) is (among all continuous functions of the sufficient statistic T) the unique unbiased estimate of θ. However, in order to apply Blackwell's theorem, we must show that $\psi(T)$ is also a sufficient statistic. From (15) we obtain

$$\frac{d\psi}{dT} = 1 + \frac{\dfrac{d^2}{dt^2}[\ln Q(T)]}{n\left[\dfrac{d}{dt}(\ln Q(T))\right]^2}.$$

From Assumption D it follows that for all sample sizes $n \geq 1$ we have $1 + \dfrac{M}{n} > \dfrac{d\psi}{dT} > 0$. Hence the function $\psi(T)$ establishes a one-to-one correspondence between T and $\psi(T)$ except possibly at a set of points of probability measure zero. Therefore $\psi(T)$ as defined in (15) is a sufficient statistic. It follows immediately from Blackwell's theorem and the existence of a unique unbiased estimate among all functions of T that $\psi(T)$ is the best estimate of the unknown parameter θ.

THEOREM 4. *If a probability distribution with range from θ to $b(\theta)$ satisfies Assumptions A and B with $K(x) \equiv 0$, and if the upper extremity of the range, $b(\theta)$, is not a strictly monotone decreasing function of θ, there exists no single sufficient statistic for θ, which is a single valued function of the values of n independent drawings from the population.*

PROOF. Under the assumptions of the Theorem to be established we have proved in Theorem 2 that $x_{(1)}$ and $x_{(n)}$ are a sufficient set of statistics for θ. We may therefore confine our attention to a search for a single valued function $T(x_{(1)}, x_{(n)})$. It is clear that

(16)
$$Pr\{t < T < t + dt\} = n(n-1)[Q(\theta)]^n \iint_{t<T<t+dt} P(x_{(1)})P(x_{(n)})$$
$$\cdot \left[\int_{x_{(1)}}^{x_{(n)}} P(\eta) \, d\eta\right]^{n-2} dx_{(1)} \, dx_{(n)}.$$

Since the likelihood function of the ensemble of n independent observations taken from the distribution has (under our assumptions as to its form) the factor $[Q(\theta)]^n$ as the sole term involving θ, it is evident from the definition of sufficiency that the integral

(17)
$$\iint_{t<T<t+dt} P(x_{(1)})P(x_{(n)})\left[\int_{x_{(1)}}^{x_{(n)}} P(\eta) \, d\eta\right]^{n-2} dx_{(1)} \, dx_{(n)},$$

when evaluated over the region common to the strip $t < T < t + dt$ and the triangle $\theta \leqq x_{(1)} \leqq x_{(n)}$, $\theta \leqq x_{(n)} \leqq b(\theta)$ in the $x_{(1)}$, $x_{(n)}$ plane must be independent of θ except in the case in which the strip includes a finite length of either the line $x_{(1)} = \theta$ or the line $x_{(n)} = b(\theta)$. Moreover this restriction must be satisfied uniformly in θ for $\theta \,\epsilon\, \Omega$. The situation is clarified by looking at Figure 2.
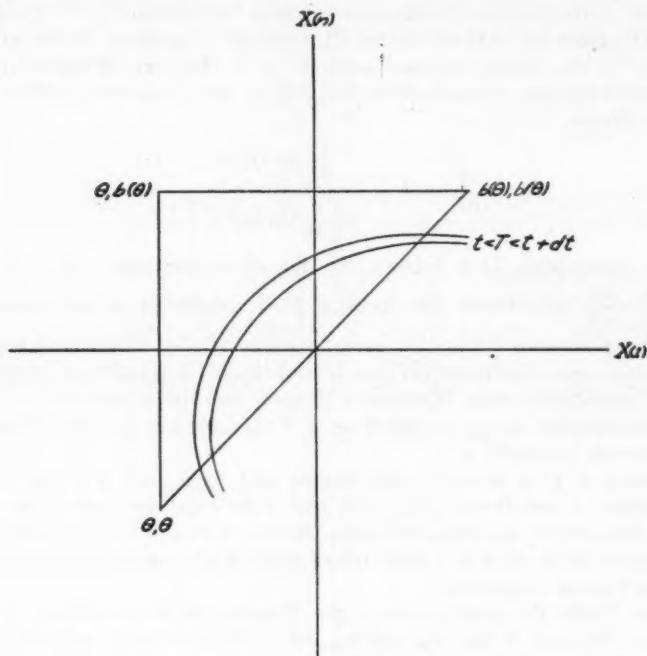

FIGURE 2

It is clear from Figure 2 that if the strip $t < T < t + dt$ does not enter and leave the triangle along the line $x_{(1)} = x_{(n)}$ without crossing either of the other two sides *for every* θ in Ω, the integral in (17) will be a function of θ. Suppose that the statistic T is of such a form that one of its strips $t < T < t + dt$ does not consist of the portions of two straight lines as was the case in Figure 1. Then for some $\theta_1 \,\epsilon\, \Omega$ this strip $t < T < t + dt$ will intersect the triangle corresponding to the value θ_1 somewhere along at least one of the lines $x_{(1)} = \theta_1$ or $x_{(n)} = b(\theta_1)$. It follows that the contours $T = $ constant must be of the same type as shown in Figure 1 *regardless* of the nature of the function $b(\theta)$.

Next we proceed to show that if $b(\theta)$ is not strictly monotone decreasing, the assumption that T is a single valued function of $x_{(1)}$ and $x_{(n)}$ is violated. The

argument proceeds as follows: under the assumptions of the theorem $b(\theta)$ is a continuous function of θ which is not strictly monotone decreasing. Hence there exist at least two values of $\theta \, \epsilon \, \Omega$, say θ_1 and θ_2, such that the corresponding contours of fixed T, say T_1 and T_2 intersect at least in one point P. The situation is shown in Figure 3. Now obviously $T_1 = T_2$, since otherwise $T(x_{(1)}, x_{(n)})$ would not be a single valued function of $x_{(1)}$ and $x_{(n)}$. From the properties of
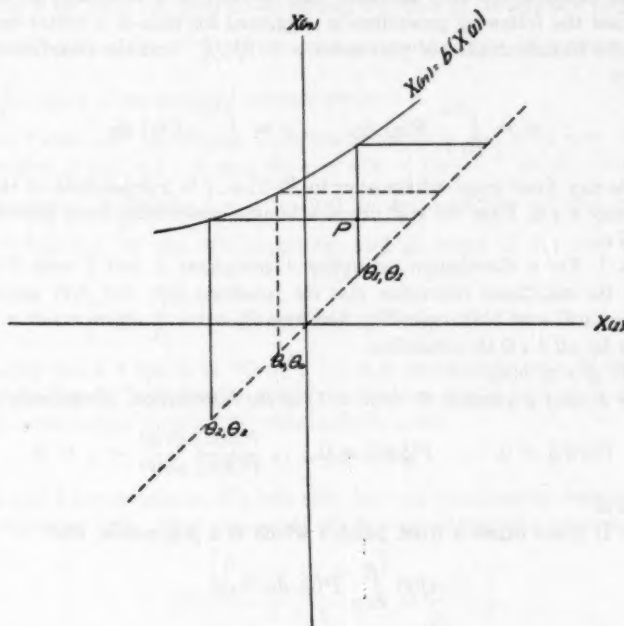
FIGURE 3

the function $b(\theta)$ there exists a $\theta_0 \, \epsilon \, \Omega$ such that the triangle defined by $\theta_0 \leqq x_{(1)} \leqq x_{(n)}$, $\theta_0 \leqq x_{(n)} \leqq b(\theta_0)$ includes a finite length of the contour $T_1 = T_2 =$ constant. Moreover since this contour cuts the above triangle at one or more points whose coordinates depend upon the value of θ_0, it follows that if the true value of θ is θ_0, the integral defined in (17) will be a function of θ_0. Hence T is not a sufficient statistic for θ for the true value lying in the parameter set Ω.

6. An alternative approach when a single sufficient statistic does not exist. It follows from Theorem 4 that if $b(\theta)$ is not a strictly decreasing monotone function of θ that no single sufficient statistic exists. The question remains as

to what to do to obtain an estimate for θ. The following procedure yields an unbiased estimate for a certain function of θ which is "best" only in the sense that it has minimum variance among the class of all analytic functions of two prescribed functions of $x_{(1)}$ and $x_{(n)}$. The fact that the sufficient statistic first derived by Pitman; i.e., $T = \min. \{x_{(1)}, b^{-1}(x_{(n)})\}$ is not an analytic function of $x_{(1)}$ and $x_{(n)}$ throughout the triangle $\theta \leqq x_{(1)} \leqq x_{(n)}$, $\theta \leqq x_{(n)} \leqq b(\theta)$ suggests that perhaps the best estimate may always be a non-analytic function. In any case the following procedure is suggested for lack of a better one.

Make the transformation of parameter $\varphi = [Q(\theta)]^{-1}$ and the coordinate transformation

$$u = \int_{x_{(1)}}^{x_{(n)}} P(\eta)\ d\eta, \qquad v = \int_{c}^{x_{(1)}} P(\eta)\ d\eta,$$

where c is any fixed point whatsoever in R_1; i.e., c is independent of the value of θ for any $\theta \in \Omega$. First we will prove a lemma concerning fixed points of the nature of c.

LEMMA 1. *For a distribution satisfying Assumptions A and B with $K(x) \equiv 0$ and with the additional restriction that the functions $a(\theta)$ and $b(\theta)$ possess first derivatives ($a(\theta)$ and $b(\theta)$ depending non-trivially upon θ), there exists a point c satisfying for all $\theta \in \Omega$ the conditions*

1.) $a(\theta) \leqq c \leqq b(\theta)$,

2.) *c is a fixed p-quantile $(0 < p < 1)$ of the distribution, if and only if*

$$P[a(\theta)] \neq 0, \qquad P[b(\theta)] \neq 0, \qquad \frac{P[b(\theta)]}{P[a(\theta)]} \frac{db(\theta)}{da(\theta)} = \rho < 0.$$

for all $\theta \in \Omega$.

PROOF. If there exists a fixed point c which is a p-quantile, the

$$(18) \qquad Q(\theta) \int_{a(\theta)}^{c} P(\eta)\ d\eta = p.$$

Writing $q(\theta) = \dfrac{1}{Q(\theta)} = \displaystyle\int_{a(\theta)}^{b(\theta)} P(\eta)\ d\eta,$

and differentiating (18) with respect to θ,

$$-\frac{da}{d\theta} P[a(\theta)] = p \frac{dq}{d\theta} = p \left\{ \frac{db}{d\theta} P[b(\theta)] - \frac{da}{d\theta} P[a(\theta)] \right\}.$$

Solving for p, we obtain

$$p = \frac{1}{1 - \dfrac{P[b(\theta)]}{P[a(\theta)]} \dfrac{db}{da}}.$$

Since there is at most one value of p obtained from (19), and since $P(x) > 0$, it follows from (18) that c is a single valued function of p. This completes the proof of the lemma.

It is clear from Lemma 1 that in the case we are now considering there exists no fixed point c which is a p-quantile of the distribution, since $\dfrac{db}{da}$ is not negative for all $\theta \, \epsilon \, \Omega$. We are now ready to prove the following theorem:

THEOREM 5. *For a distribution satisfying Assumptions A and B with $K(x) \equiv 0$ and with the additional restriction that $b(\theta)$ is not a strictly monotone decreasing function of θ for all $\theta \, \epsilon \, \Omega$, there exists among the class of all analytic functions of* $u = \displaystyle\int_{x_{(1)}}^{x_{(n)}} P(\eta) \, d\eta$ *and* $v = \displaystyle\int_{c}^{x_{(1)}} P(\eta) \, d\eta$ *a unique function of u and v; namely* $\left(\dfrac{n+1}{n-1}\right) u$, *which is an unbiased estimate for φ.*

PROOF. Under our coordinate transformation to u and v as new variables of integration, $g(x_{(1)} , x_{(n)} ; \varphi) \, dx_{(1)} \, dx_{(n)} = n(n-1)\varphi^{-n} u^{n-2} \, du \, dv$. Introducing a new function of θ; namely, $\beta = \displaystyle\int_{c}^{a(\theta)} P(\eta) \, d\eta$ the condition (11) for unbiasedness in θ becomes for the new parameter and in terms of the new variables u and v,

$$(20) \qquad \int_{0}^{\varphi} du \int_{\beta}^{\varphi - u + \beta} n(n-1)\varphi^{-n} u^{n-2} \psi(u, v) \, dv \equiv \varphi$$

for all φ for which θ lies in Ω, where $\psi(u, v)$ is an estimate of φ. If we expand $\psi(u, v)$ in a double Taylor series about the point $u = 0$, $v = 0$, it is clear that the only terms which satisfy (20) identically in φ are

$$\psi(u, v) = au + bv,$$

where a and b are constants. We will now derive a relationship between a and b by integrating (20). After some easy algebra we obtain the relationship

$$(21) \qquad a + b \dfrac{\left[\dfrac{\beta}{\varphi}(n+1) + 1 \right]}{n-1} = \dfrac{n+1}{n-1}.$$

Under the conditions of the Theorem it is clear from Lemma 1 that the point c is not a p-quantile uniformly in φ and hence $\dfrac{\beta}{\varphi}$ is not a constant independent of φ. Hence the only solution of (21) is given by $a = \dfrac{n+1}{n-1}$, $b = 0$; and the only unbiased estimate of φ is

$$(22) \qquad \psi = \dfrac{n+1}{n-1} \int_{x_{(1)}}^{x_{(n)}} P(\eta) \, d\eta.$$

REFERENCES

[1] E. J. G. PITMAN, "Sufficient statistics and intrinsic accuracy," *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 567–579.

[2] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Am. Math. Soc.*, Vol. 39 (1936), pp. 399–409.

[3] DAVID BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

ON THE DISTRIBUTION OF WALD'S CLASSIFICATION
STATISTIC

By Harman Leon Harter

Michigan State College

Summary. In this paper we shall consider the exact distribution of Wald's classification statistic V in the univariate case, some theoretical approximations in various multivariate cases, and an empirical distribution in a particular multivariate case. We shall also draw some conclusions as to the potential usefulness of the statistic V and the work which remains to be done.

1. Introduction. In many educational and industrial problems it is necessary to classify persons or objects into one of two categories—those fit and those unfit for a particular purpose. In formulating this problem of classification, Wald [1] assumed that for p tests we know the scores of N_1 individuals known to belong to population Π_1 and of N_2 individuals known to belong to population Π_2, along with those of the individual under consideration, a member of the population Π, where it is known a priori that Π is identical with either Π_1 or Π_2. He assumed moreover that the distribution of the test scores of the individuals making up Π_1 and Π_2 are two p-variate normal distributions which have the same covariance matrix, but are independent of each other. In order to classify the individual in question into either Π_1 or Π_2, Wald introduced the statistic V defined by the relation

$$(1) \qquad V = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} t_{i,n+1} t_{j,n+2} \qquad (n = N_1 + N_2 - 2),$$

where

$$(2) \qquad \| s^{ij} \| = \| s_{ij} \|^{-1}, \quad s_{ij} = \frac{\sum_{\alpha=1}^{n} t_{i\alpha} t_{j\alpha}}{n},$$

and where the variates $t_{i\beta}$ ($i = 1, \cdots, p; \beta = 1, \cdots, n + 2$) are normally and independently distributed with unit variance and with expected values

$$(3) \qquad E(t_{i\alpha}) = 0 \ (\alpha = 1, \cdots, n), \qquad E(t_{i,n+1}) = \rho_i, \qquad E(t_{i,n+2}) = \zeta_i,$$

where ρ_i and ζ_i are constants.

2. The exact distribution of V when $p = 1$. In the univariate case, the definition (1) reduces to

$$(4) \qquad V = s^{11} t_{1,n+1} t_{1,n+2},$$

where

$$s^{11} = \frac{1}{s_{11}} = \frac{1}{\sum_{\alpha=1}^{n} t_{1\alpha}^2 / n}, \qquad n = N_1 + N_2 - 2.$$

58

Thus, in the case $p = 1$,

$$(5) \qquad V = \frac{t_{1,n+1} \, t_{1n,+2}}{\sum_{a=1}^{n} t_{1a}^2 / n} = \frac{xy}{z},$$

where

$$x = t_{1,n+1}, \qquad y = t_{1,n+2}, \qquad z = \sum_{a=1}^{n} t_{1a}^2 / n.$$

In the degenerate case ($\rho_1 = \zeta_1 = 0$), x and y are normally distributed with zero means, so that their probability laws are

$$(6) \qquad P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad P(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}.$$

Because of symmetry we have then

$$(7) \qquad P(|\,x\,|) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad P(|\,y\,|) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}.$$

It is well known that $z = \sum_{a=1}^{n} t_{1a}^2 / n$ is distributed as χ^2 / n with n degrees of freedom, that is, the probability law for z is

$$(8) \qquad P(z) = \frac{n^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \frac{z^{\frac{1}{2}n-1} e^{-\frac{1}{2}nz}}{2^{\frac{1}{2}n}}.$$

Now we proceed to find the probability law of $V = \dfrac{|\,x\,| \cdot |\,y\,|}{z}$ in a manner similar to that used by Shrivastava [2] in investigating a different statistic. Let $w = \ln |\,V\,| = \ln |\,x\,| + \ln |\,y\,| - \ln z$. Then the characteristic function of w is given by

$$(9) \qquad \phi(t) = \int_0^{\infty}\!\!\int_0^{\infty}\!\!\int_0^{\infty} e^{iwt} P(|\,x\,|) P(|\,y\,|) P(z) \, dx \, dy \, dz.$$

Substituting the values of $P(|\,x\,|)$, $P(|\,y\,|)$ and $P(z)$ from (7) and (8), and making use of the independence of x, y and z, we have

$$(10) \qquad \phi(t) = \frac{n^{\frac{1}{2}n}}{2^{\frac{1}{2}n-1} \Gamma(\frac{1}{2}n)\pi} \int_0^{\infty} x^{it} e^{-\frac{1}{2}x^2} \, dx \int_0^{\infty} y^{it} e^{-\frac{1}{2}y^2} \, dy \int_0^{\infty} z^{\frac{1}{2}n-1-it} e^{-\frac{1}{2}nz} \, dz.$$

Expressing the integrals in (10) in terms of Gamma functions and simplifying, we find

$$(11) \qquad \phi(t) = \frac{n^{it}}{\pi \Gamma(\frac{1}{2}n)} \Gamma(\tfrac{1}{2}n - it) \left[\Gamma\left(\frac{it+1}{2}\right) \right]^2.$$

Upon inserting this result in the Levy inversion formula

$$(12) \qquad P(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwt} \phi(t) \, dt$$

and making the substitution $v = it$, we obtain

$$(13) \qquad P(w) = \frac{n^{it}}{2\pi^2 i \Gamma(\frac{1}{2}n)} \int_{-i\infty}^{+i\infty} e^{-vw} \Gamma(\tfrac{1}{2}n - v) \left[\Gamma\left(\frac{v+1}{2}\right) \right]^2 dv.$$

Using a property of the Gamma function given by Whittaker and Watson [3]

$$(14) \qquad \Gamma(z)\Gamma(1 - z) = \frac{\pi}{\sin \pi z}$$

and letting $z = \frac{1}{2}n - v$, we obtain

$$(15) \qquad \Gamma(\tfrac{1}{2}n - v) = \frac{\pi}{\Gamma(v - \frac{1}{2}n + 1) \sin \pi(\frac{1}{2}n - v)}.$$

Substituting this value of $\Gamma(\frac{1}{2}n - v)$ in (13), and simplifying, we find

$$(16) \qquad P(w) = \frac{1}{\Gamma(\frac{1}{2}n)} \cdot \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} e^{-vw} \frac{n^v \left[\Gamma\left(\dfrac{v+1}{2}\right) \right]^2}{\sin \pi(\frac{1}{2}n - v)\Gamma(v - \frac{1}{2}n + 1)} \, dv.$$

We shall now perform a contour integration, using as the contour the imaginary axis plus the semicircle in the right half-plane with center at the origin and infinite radius. It can be shown that for $|\, n/2e^w\,| < 1$, and hence for $|\, V\,| > n/2$, the integral around the semicircular portion of the contour is zero. Hence, under these conditions, the integral on the right side of (16) is equal to $(-2\pi i)$ times the sum of the residues at all the singular points in the right half-plane. The integrand has simple poles at $v = \frac{1}{2}n, \frac{1}{2}n + 1, \frac{1}{2}n + 2, \cdots$, and no other singularities in the right half-plane. Inserting the actual values of the residues, using the fact that $\cos k\pi = (-1)^k$, for k an integer, and letting $v = j + \frac{1}{2}n$, we find

$$(17) \qquad P(w) = \frac{1}{\pi \Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} e^{-(j+\frac{1}{2}n)w} n^{j+\frac{1}{2}n} (-1)^j \frac{\left[\Gamma\left(\dfrac{2j+n+2}{4}\right) \right]^2}{\Gamma(j+1)}.$$

Replacing e^w by $|\, V\,|$ and multiplying by $\dfrac{dw}{d\,|\,V\,|} = \dfrac{1}{|\,V\,|}$, we obtain the probability law for $|\,V\,|$

$$(18) \qquad P(|\, V\,|) = \frac{1}{\pi \Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} n^{j+\frac{1}{2}n} |\, V\,|^{-j-\frac{1}{2}n-1} (-1)^j \frac{\left[\Gamma\left(\dfrac{2j+n+2}{4}\right) \right]^2}{\Gamma(j+1)}.$$

The infinite series on the right side of (18) converges for precisely those values of $|\, V\,|$ for which the integral along the semicircular portion of the path is zero, that is for $|\, V\,| > \frac{1}{2}n$. Since the values of x and y are symmetric about zero and uncorrelated, the values of V are also symmetric about zero, and hence $P(V) = \frac{1}{2}P(|\, V\,|)$.

To obtain a series for $P(\mid V \mid)$ which converges when $\mid V \mid < \frac{1}{2}n$, it would be necessary to perform a contour integration around the left half-plane, which is considerably more difficult, since the presence of $\left[\Gamma \left(\frac{v+1}{2} \right) \right]^2$ in the integrand of (16) introduces double poles at $v = -1, -3, -5, \cdots$.

If we drop the restriction $\zeta_1 = 0$, but keep $\rho_1 = 0$, V will still be distributed symmetrically about zero, since x is distributed symmetrically about zero and is independent of y. The probability laws for x and z will be the same as in the degenerate case, but $P(\mid V \mid)$ will be different, due to a change in $P(\mid y \mid)$. Since the mean of the distribution of y's is now $\zeta_1 \neq 0$, we have

$$(19) \qquad P(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\zeta_1)^2},$$

which yields

$$(20) \quad P(\mid y \mid) = \frac{1}{\sqrt{2\pi}} [e^{-\frac{1}{2}(y-\zeta_1)^2} + e^{-\frac{1}{2}(-y-\zeta_1)^2}] = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta_1^2} e^{-\frac{1}{2}y^2} \sum_{r=0}^{\infty} \frac{(y\zeta_1)^{2r}}{(2r)!}.$$

Proceeding in the same manner as for the degenerate case, we find as the characteristic function of $w = \ln \mid V \mid$

$$(21) \quad \phi(t) = \frac{n^{it} e^{-\frac{1}{2}\zeta_1^2}}{\pi \Gamma(\frac{1}{2}n)} \Gamma\left(\frac{it+1}{2}\right) \Gamma(\frac{1}{2}n - it) \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{it+1}{2}\right).$$

Again using the Levy inversion formula (12), and letting $v = it$, we have

$$(22) \qquad \begin{aligned} P(w) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{2\pi^2 i \Gamma(\frac{1}{2}n)} \int_{-i\infty}^{+i\infty} n^v e^{-vw} \Gamma\left(\frac{v+1}{2}\right) \\ \cdot \Gamma(\frac{1}{2}n - v) \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right) dv. \end{aligned}$$

This integral may be evaluated by integrating around the same contour as in the degenerate case. Performing the contour integration and simplifying, we obtain

$$(23) \quad P(w) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi \Gamma(\frac{1}{2}n)} \sum_{v=-\frac{1}{2}n}^{\infty} \frac{n^v e^{-vw} \Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v - \frac{1}{2}n + 1)} (-1)^{v+\frac{1}{2}n} \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right).$$

Replacing e^w by $\mid V \mid$ and multiplying by $\dfrac{dw}{d \mid V \mid} = \dfrac{1}{\mid V \mid}$, we find

$$(24) \quad P(\mid V \mid) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi \Gamma(\frac{1}{2}n)} \sum_{v=-\frac{1}{2}n}^{\infty} \frac{n^v \Gamma\left(\frac{v+1}{2}\right) (-1)^{v+\frac{1}{2}n}}{\mid V \mid^{v+1} \Gamma(v - \frac{1}{2}n + 1)} \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right).$$

Letting $v = j + \frac{1}{2}n$, this may be written in the form

$$(25) \quad P(|V|) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi\Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} \frac{(-1)^j n^{j+\frac{1}{2}n} \Gamma\left(\frac{2j+n+2}{4}\right)}{|V|^{j+\frac{1}{2}n+1} \Gamma(j+1)}$$
$$\cdot \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{2j+n+2}{4}\right).$$

This expression is valid (since the integral vanishes along the semicircular portion of the contour) and converges for precisely the same values of V as in the degenerate case, that is for $|V| > n/2$.

3. Approximate distributions of V in various multivariate cases. Wald [1] has shown that the distribution of the statistic V is the same as that of the statistic

$$(26) \quad \bar{V} = -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)},$$

where the joint distribution of m_1, m_2 and m_3 is known. Since m_1, m_2 and m_3 are of the order $1/n$ in the probability sense, the denominator of (26) is near -1 nearly always for sufficiently large n. Accordingly, Wald has suggested that even for moderately large n, V is distributed approximately as nm_3. By integrating out m_1 and m_2 over the domain for which the joint distribution is real and ≥ 0, it is possible to find the distribution of m_3, and from it the distribution of nm_3, which is approximately the distribution of V, for sufficiently large n. We restrict ourselves to values of n and p satisfying the relation $1 < p \leq n$. Four cases have to be considered: (1a) n even, p odd; (1b) n even, p even; (2a) n odd, p even; and (2b) n odd, p odd.

For the degenerate case $\rho_i = \zeta_i = 0$, it can be shown that the joint distribution of m_1, m_2 and m_3 given by Wald [1] reduces to

$$(27) \quad C[(1 - m_1)(1 - m_2) - m_3^2]^{(n-1-p)/2}[m_1 m_2 - m_3^2]^{(p-3)/2} dm_1 dm_2 \, dm_3 \,,$$

where C is a constant. In integrating out m_1 and m_2, we must be careful to integrate over only the domain for which the joint distribution (27) is real and ≥ 0. This requires that the following inequalities hold:

$$(28) \quad m_1 m_2 - m_3^2 \geq 0, \quad (1 - m_1)(1 - m_2) - m_3^2 \geq 0.$$

From these it follows that the limits for m_1 and m_2 are

$$(29) \quad \frac{m_3^2}{m_2} \leq m_1 \leq 1 - \frac{m_3^2}{1 - m_2}, \quad \frac{1 - \sqrt{1 - 4m_3^2}}{2} \leq m_2 \leq \frac{1 + \sqrt{1 - 4m_3^2}}{2}.$$

For Case 1a (n even, p odd), let $p = 3 + 2c$, where $c =$ an integer ≥ 0. The distribution function $G_{n,p}(m_3)$ can then be expressed as a double integral, as follows:

$$(30) \quad G_{n,3+2c}(m_3) = C \int_{(1-\sqrt{1-4m_3^2})/2}^{(1+\sqrt{1-4m_3^2})/2} \int_{m_3^2/m_2}^{1-m_3^2/(1-m_2)} [(1-m_1)(1-m_2)-m_3^2]^{(n-4)/2-c}$$
$$\cdot [m_1 m_2 - m_3^2]^c \, dm_1 dm_2.$$

Expanding repeatedly by the binomial theorem and integrating out m_1, then expanding again and integrating out m_2, we find

$$G_{n,3+2c}(m_3) = C \sum_{j=0}^{(n-4)/2-c} \binom{\frac{n-4}{2}-c}{j}$$

$$(31) \qquad \cdot \sum_{k=0}^{j} \binom{j}{k} \sum_{q=0}^{c} (-1)^{j+q} \binom{c}{q} \frac{2}{n-2-2j-2q}$$

$$[A_{j,k,q}(m_3) + B_{j,k,q}(m_3) - C_{j,k}(m_3) - D_{j,k}(m_3)],$$

where

$$A_{j,k,q}(m_3) = \sum_{r=0}^{\min[(n-2)/2-j-q,(n-4)/2-c-k]} \binom{\frac{n-2}{2}-j-q}{r} m_3^{2(k+q+r)}$$

$$(32) \qquad \cdot \sum_{t=0}^{(n-4)/2-c-k-r} (-1)^t \binom{\frac{n-4}{2}-c-k-r}{t} \frac{2}{n-2-2k-2q-2r-2t}$$

$$\cdot \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r-t} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r-t} \right],$$

$$B_{j,k,q}(m_3) = \sum_{r'=(n-2)/2-c-k}^{(n-2)/2-j-q} \binom{\frac{n-2}{2}-j-q}{r'} m_3^{2(k+q+r')}$$

$$(33) \qquad \cdot \left\{ \binom{c-q}{r'+c+k-\frac{n-2}{2}} \ln \frac{1-\sqrt{1-4m_3^2}}{1+\sqrt{1-4m_3^2}} \right.$$

$$+ \sum_{\substack{t'=0 \\ t' \neq r'+c+k-(n-2)/2}}^{c-q} (-1)^{t'-r'-c-k+(n-2)/2} \binom{c-q}{t'} \frac{2}{2t'-2r'-2c-2k+n-2}$$

$$\cdot \left. \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r'-t'} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r'-t'} \right] \right\},$$

$$C_{j,k}(m_3) = \sum_{\substack{s=0 \\ s \neq j-k}}^{(n-4)/2-c-k} (-1)^s \binom{\frac{n-4}{2}-c-k}{s} \frac{1}{j-k-s} m_3^{n-2+2k-2j}$$

$$(34) \qquad \cdot \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{j-k-s} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{j-k-s} \right],$$

$$(35) \quad D_{j,k}(m_3) = (-1)^{j-k+1} \binom{\dfrac{n-4}{2}-c-k}{j-k} m_3^{n-2+2k-2j} \ln \frac{1-\sqrt{1-4m_3^2}}{1+\sqrt{1-4m_3^2}},$$

the terms involving natural logarithms having the value zero when $m_3 = 0$. As a numerical example we have, after normalization,

$$(36) \quad G_{10,2}(m_3) = \frac{180}{\pi}\left[\left(\tfrac{1}{16} + \tfrac{5}{24}m_3^2 + \tfrac{1}{4}m_3^4 + \tfrac{1}{4}m_3^6\right)\sqrt{1-4m_3^2} \right.$$

$$\left. - \left(m_3^2 + \tfrac{3}{2}m_3^4 - \tfrac{1}{2}m_3^6\right)\ln\frac{1+\sqrt{1-4m_3^2}}{1-\sqrt{1-4m_3^2}}\right].$$

For Case 1b (n even, p even), let $p = 2 + 2c$, where $c =$ an integer ≥ 0. The distribution function $G_{n,p}(m_3)$ can then be expressed as a double integral, as follows:

$$(37) \quad G_{n,2+2c}(m_3) = C \int_{(1-\sqrt{1-4m_3^2})/2}^{(1+\sqrt{1-4m_3^2})/2} \int_{m_3^2/m_2}^{1-m_3^2/(1-m_2)} [(1-m_1)(1-m_2)-m_3^2]^{(n-3)/2-c}$$

$$\cdot [m_1 m_2 - m_3^2]^{c-1}\, dm_1\, dm_2.$$

This double integration can be performed by the use of certain formulas given by Peirce [4], and after evaluation we have

$$G_{n,2+2c}(m_3) = C\cdot 2\pi(-1)^{(n-2)/2-c}$$

$$\cdot \frac{(2c-1)(2c-3)\cdots 1(n-2c-3)(n-2c-5)\cdots 1}{(n-2)(n-4)\cdots 2}$$

$$(38) \quad \cdot \left[\sum_{\substack{j=0 \\ (j-k)\leq c}}^{(n-2)/2} \sum_{k=0}^{(n-2)/2-j} (-1)^{(n-2)/2+k-c-j} \binom{\dfrac{n-2}{2}}{j}\binom{\dfrac{n-2}{2}-j}{k} A'_{j,k}(m_3) \right.$$

$$\left. + \sum_{\substack{j=0 \\ (j-k)>c}}^{(n-2)/2} \sum_{k=0}^{(n-2)/2-j} (-1)^{(n-2)/2+k-c-j} \binom{\dfrac{n-2}{2}}{j}\binom{\dfrac{n-2}{2}-j}{k} B'_{j,k}(m_3) \right],$$

where

$$A'_{j,k}(m_3) = m_3^{2j} \left[m_3 \sum_{q=0}^{c} (-1)^q \right.$$

$$\cdot \frac{(2m - 2c + 3)(2m - 2c + 1) \cdots (2m - 2c - 2q + 5)}{(2c - 1)(2c - 3) \cdots (2c - 2q - 1)}$$

$$\cdot \left\{ \frac{\left(\dfrac{1 + \sqrt{1 - 4m_3^2}}{2} \right)^{m+1}}{\left(\dfrac{1 - \sqrt{1 - 4m_3^2}}{2} \right)^{c-q}} - \frac{\left(\dfrac{1 - \sqrt{1 - 4m_3^2}}{2} \right)^{m+1}}{\left(\dfrac{1 + \sqrt{1 - 4m_3^2}}{2} \right)^{c-q}} \right\}$$

(39)

$$+ (-1)^c \frac{(2m - 2c + 3)(2m - 2c + 1) \cdots (2m + 3)}{(2c - 1)(2c - 3) \cdots 1}$$

$$\cdot \left((-1)^{m+1} \frac{2m(2m - 2) \cdots 1}{(2m + 1)(2m - 1) \cdots 2} (-\sin^{-1} \sqrt{1 - 4m_3^2}) \right.$$

$$- m_3 \sum_{r=0}^{m-1} \frac{2m(2m - 2) \cdots (2m - 2r + 2)}{(2m + 1)(2m - 1) \cdots (2m - 2r + 1)}$$

$$\left. \left. \cdot \left\{ \left(\frac{1 + \sqrt{1 - 4m_3^2}}{2} \right)^{m-r-1} - \left(\frac{1 - \sqrt{1 - 4m_3^2}}{2} \right)^{m-r-1} \right\} \right) \right],$$

$$B'_{j,k}(m_3) = m_3^{2j} \left[m_3 \sum_{q'=0}^{c} \right.$$

$$\cdot \frac{(2m' + 2c - 3)(2m' + 2c - 5) \cdots (2m' + 2c - 2q' - 1)}{(2c - 1)(2c - 3) \cdots (2c - 2q' - 1)}$$

$$\cdot \left\{ \frac{\left(\dfrac{2}{1 + \sqrt{1 - 4m_3^2}} \right)^{m'-1}}{\left(\dfrac{2}{1 - \sqrt{1 - 4m_3^2}} \right)^{c-q'}} - \frac{\left(\dfrac{2}{1 - \sqrt{1 - 4m_3^2}} \right)^{m'-1}}{\left(\dfrac{1 + \sqrt{1 - 4m_3^2}}{2} \right)^{c-q'}} \right\}$$

(40)

$$+ \frac{(2m' + 2c - 3)(2m' + 2c - 5) \cdots (2m' - 3)}{(2c - 1)(2c - 3) \cdots 1}$$

$$\cdot m_3 \sum_{r'=0}^{m'-1} (-1)^{r'+1} \frac{(2m' - 3)(2m' - 5) \cdots (2m' - 2r' - 1)}{(2m' - 2)(2m' - 4) \cdots (2m' - 2r' - 2)}$$

$$\left. \cdot \left\{ \left(\frac{2}{1 + \sqrt{1 - 4m_3^2}} \right)^{m'-r'-1} - \left(\frac{2}{1 - \sqrt{1 - 4m_3^2}} \right)^{m'-r'-1} \right\} \right],$$

$$(41) \qquad m = k + c - j - \tfrac{1}{2}, \qquad m' = j - k - c + \tfrac{1}{2}.$$

As a numerical illustration we have, after normalization,

$$G_{10,2}(m_3) = \left(\frac{55125}{16384} + \frac{23625}{256} m_3^2 + \frac{4725}{64} m_3^4\right) \sin^{-1}\sqrt{1 - 4m_3^2}$$

(42)

$$- \left(\frac{313515}{8192} + \frac{99825}{1024} m_3^2 - \frac{465}{64} m_3^4 - \frac{45}{8} m_3^6\right)|m_3|\sqrt{1 - 4m}.$$

In Cases 2a and 2b, infinite series of elliptic integrals occur, and it appears that approximate integration is the best than can be done.

The author plans a later paper on the distribution for the nondegenerate case $\rho_i = 0$, $\zeta_i \neq 0$.

For small values of n, Wald's approximation nm_3 is not applicable. One can obtain a fair approximation by replacing $1/[m_3^2 - (1 - m_1)(1 - m_2)]$ in (26) by its average with respect to m_1 and m_2 over the domain, taking account of the joint distribution function (27). This yields

(43)
$$V \doteq \frac{C_{n,p}}{C_{n-2,p}} nm_3 \frac{G_{n-2,p}(m_3)}{G_{n,p}(m_3)},$$

where $C_{n-2,p}$ and $C_{n,p}$ are the constants in the joint distribution of m_1, m_2 and m_3 for the values of n and p involved. The approximation (43), while rather crude, is better than Wald's nm_3 for small values of n, and asymptotically equivalent to it as $n \to \infty$.

4. An empirical distribution of V. A sampling experiment was performed in order to obtain an empirical distribution of 1000 values of V for $n = 10$, $p = 3$, $\rho_i = \zeta_i = 0$. Ten thousand wooden beads were stamped with two digit numbers whose distribution approximates as nearly as possible that of a normal population with mean 50 and standard deviation 10. One thousand sets of values $x_{i\alpha}(i = 1, 2, 3; \alpha = 1, 2, \cdots, 12)$ were obtained by sampling with replacement from this population. The values $x_{i\alpha}$ were expressed in standard units $t_{i\alpha}$, using

(44)
$$t_{i\alpha} = \frac{x_{i\alpha} - 50}{10}.$$

From the standard variables $t_{i\alpha}$, one thousand values of V were calculated by means of (1) and (2), using IBM equipment. The resulting empirical distribution is given in Table 1. This distribution was compared with the theoretical approximation (43), which is, for $n = 10$, $p = 3$, $\rho_i = \zeta_i = 0$

(45)
$$V \doteq \frac{150}{7} m_3 \frac{G_{8,3}(m_3)}{G_{10,3}(m_3)}.$$

The approximation fits the observed distribution fairly well for the central classes, but underestimates the frequencies of large values of $|V|$ quite badly.

5. Conclusions. The statistic V is potentially very useful, but much work remains to be done in obtaining the necessary information about its distribution, especially in the small sampling case, and tabulating the associated probabilities. Even in the univariate case, where the exact distribution is known, the amount

of labor involved in determining probabilities is very great and a simple approximation is needed, unless a high speed computing device is available. For the multivariate small sampling case, only a crude approximation to the distribution of V is available, and the exact distribution or a better approximation is needed.

TABLE 1

Frequency distribution of 1000 empirical values of V for $n = 10$,
$p = 3$, $\rho_i = \zeta_i = 0$ (Class marks integers)

Class mark	Frequency f	Class mark	Frequency f	Class mark	Frequency f
76	1	12	3	−8	15
...		11	3	−9	12
44	1	10	3	−10	6
...		9	6	−11	3
39	1	8	10	−12	2
...		7	16	−13	2
30	1	6	11	−14	3
29	1	5	15	−15	4
...		4	33	...	
24	1	3	54	−18	2
23	1	2	85	...	
...		1	140	−23	1
20	1	0	181	−24	2
19	2	−1	134	...	
18	1	−2	101	−28	1
...		−3	52	−29	1
16	1	−4	26	...	
15	2	−5	17	−36	1
14	1	−6	23	−37	1
13	4	−7	12		

$$\bar{V} = -.0700, \qquad \sigma_V = 5.938$$

REFERENCES

[1] A. WALD, "On a statistical problem arising in the classification of an individual into one of two groups," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[2] M. P. SHRIVASTAVA, "On the D^2-statistic," *Bull. Calcutta Math. Soc.*, Vol. 33 (1941), pp. 71–86.

[3] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, 4th ed., Cambridge University Press, London, 1940, p. 239.

[4] B. O. PEIRCE, *A Short Table of Integrals*, Ginn and Co., Boston, 1929, pp. 18–19, Formulas 113, 118, 119, 120, 121, 123.

RATIOS INVOLVING EXTREME VALUES[1]

By W. J. Dixon

University of Oregon

1. Summary. Ratios of the form $(x_n - x_{n-j})/(x_n - x_i)$ for small values of i and j and $n = 3, \cdots, 30$ are discussed. The variables concerned are order statistics, i.e., sample values such that $x_1 < x_2 < \cdots < x_n$. Analytic results are obtained for the distributions of these ratios for several small values of n and percentage values are tabled for these distributions for samples of size $n \leq 30$.

2. Introduction. There has been interest in the problem of gross errors in data since Chauvenet presented his solution for the problem about 1850. His hypothesis was essentially that in some samples a small portion of the observations were from a population with a different mean value. There has been research from that time up to the present on procedures suitable for treating such data.

If it is assumed that a certain percentage of "gross errors" may occur, then there are two general procedures for treating such data:

(1) A statistical treatment may be given to the data which gives very little weight to such aberrant values as may occur.

(2) A statistical test may be constructed which will indicate such values so that they may be rejected.

The functions to be discussed here were designed for testing the consistency of suspected values with the sample as a whole. Investigation of the performance of these criteria is given in another paper.

3. Critical values for r_{10}. The first statistic to be considered is

$$r_{10} = (x_n - x_{n-1})/(x_n - x_1),$$

where the subscripts on the x's indicate ordered values such that $x_1 < x_2 < \cdots < x_n$. The density function for x_1, x_{n-1}, x_n is

$$(1) \qquad \frac{n!}{(n-3)!} f(x_1) \, dx_1 \left(\int_{x_1}^{x_{n-1}} f(t) \, dt \right)^{n-3} f(x_{n-1}) \, dx_{n-1} f(x_n) \, dx_n.$$

Setting $v = x_n - x_1$, $rv = x_n - x_{n-1}$, $x = x_n$, and integrating x and v over their range of definition we have the density function of r_{10} for a sample of size n. (The subscripts on the r's will be dropped when there is no ambiguity.) This function appears as

$$(2) \qquad \frac{n!}{(n-3)!} \int_{-\infty}^{\infty} \int_{0}^{\infty} \left(\int_{x-v}^{x-r_{10}v} f(t) \, dt \right)^{n-3} f(x - v)f(x - r_{10}v)f(x)v \, dv \, dx.$$

There will be no loss in generality by considering the values x_i to have been drawn from a distribution with zero mean and unit variance, since the statistic is the ratio of two differences. It should also be noted that for symmetric populations, the distribution of $(x_n - x_{n-1})/(x_n - x_1)$ will be the same as that of $(x_2 - x_1)/(x_n - x_1)$. For the rectangular distribution the density function is

$$(3) \qquad (n - 2)(1 - r_{10})^{n-3} \qquad (0 < r_{10} < 1),$$

and the cdf is

$$(4) \qquad 1 - (1 - R_{10})^{n-2}.$$

If we set this expression equal to $1 - \alpha$ we obtain critical values of R_{10}

$$(5) \qquad R_{10\alpha} = 1 - \alpha^{1/n-2}.$$

For the more interesting case of the normal distribution, the operations indicated above are much more arduous.

$n = 3$, *Normal population*. The integral in (2) above can be evaluated to obtain the density function of r_{10} for the assumption of normality

$$(6) \qquad g_3(r_{10}) = \frac{3\sqrt{3}}{2\pi} \frac{1}{r^2 - r + 1}.$$

The integration of this density results in the cdf

$$(7) \qquad \frac{3}{\pi} \arctan \frac{2}{\sqrt{3}}\left(R_{10} - \tfrac{1}{2}\right) + \tfrac{1}{2}.$$

Upon setting this last expression equal to $1 - \alpha$, we obtain

$$(8) \qquad R_{10\alpha} = \tfrac{1}{2} + \frac{\sqrt{3}}{2} \tan \frac{\pi}{3}\left(\tfrac{1}{2} - \alpha\right).$$

$n = 4$, *Normal population*. The density function in this case becomes

$$(9) \qquad g_4(r_{10}) = \frac{3}{\pi} \frac{1}{r^2 - r + 1}\left[\frac{1 - 2r}{\sqrt{4r^2 - 4r + 3}} - \frac{r - 2}{\sqrt{3r^2 - 4r + 4}}\right].$$

If we now set the cdf equal to $1 - \alpha$ we obtain

$$(10) \qquad 5 - \frac{6}{\pi}\left[\arctan\sqrt{4R^2 - 4R + 3} + \arctan\frac{1}{R}\sqrt{3R^2 - 4R + 4}\right] = 1 - \alpha,$$

which may be written as follows by taking the tangent of both sides of this equation:

$$(11) \qquad \frac{\sqrt{4R^2 - 4R + 3} + \frac{1}{R}\sqrt{3R^2 - 4R + 4}}{1 - \frac{1}{R}\sqrt{(4R^2 - 4R + 3)(3R^2 - 4R + 4)}} = \tan\frac{\pi}{6}(\alpha + 4).$$

The integration of $g_4(r_{10})$ was performed for the first term by substituting $r = \frac{1}{2} + (1/\sqrt{2})\sqrt{x^2 - 1}$. The second term of $g_4(r_{10})$ is identical with the first if one substitutes $s = 1/r$.

$n = 5$, *Normal population*. For this case it can be shown that the density function has the following form

$$(12) \qquad g_5(r_{10}) = \frac{15\left[h(r) + h\left(\dfrac{1}{r}\right) \right]}{\pi^2(r^2 - r + 1)},$$

where

$$h(r) = \frac{2 - r}{\sqrt{3r^2 - 4r + 4}} \tan^{-1} \frac{(1 - r)\sqrt{5(3r^2 - 4r + 4)}}{3r^2 - 3r + 4}.$$

The cdf for $n = 5$ has not been obtained in a comparable form to those obtained for $n = 3, 4$. No such expressions were obtained for larger values of n. Various percentage values were computed from the above distributions and are presented in Table I. The percentage values were also obtained by numerical integrations for $n = 5, 7, 10, 15, 20, 25, 30$. Values for other values of n were obtained by interpolation. These percentage values can be obtained by a double quadrature since

$$(13) \qquad \begin{aligned} G(R_{10}) &= \int_0^R \int_{-\infty}^\infty \int_0^\infty g(r, x, v) \, dv \, dx \, dr_{10} = \\ &1 - n(n - 1) \int_0^R \int_{-\infty}^\infty \int_0^\infty \left(\int_{x-v}^{x-r_{10}v} f(t) \, dt \right)^{n-2} f(x)f(x - v) \, dv \, dx \, dr_{10}. \end{aligned}$$

This integral was evaluated for all combinations of the values of n indicated above and for $R_{10} = 0, .06, .10, .16, .21, .26, .30, .34, .40, .44, .48, .53, .56, .60, .80, .90$. These values are not regularly spaced since several computations were made before it was possible to select the particular values of R which would be most useful for evaluating $G(R_{10})$. The values of the integral in (13) were used as the base for computations for all the tables included in this paper.

4. Distribution of other ratios. It can be suggested that a ratio to test whether x_n is significantly far from x_{n-1} should avoid x_1. Let us consider $r_{11} = (x_n - x_{n-1})/(x_n - x_2)$. Its cdf is

$$(14) \qquad \int_{-\infty}^\infty \int_0^\infty \frac{n!}{(n - 2)!} \int_{-\infty}^x f(t) \, dt \left(\int_{x-v}^{x-r_{11}v} f(s) \, ds \right)^{n-3} f(x - v)f(x) \, dv \, dx.$$

For the rectangular distribution we obtain the density function

$$(15) \qquad (n - 3)(1 - r_{11})^{n-4}.$$

For the rectangular distribution we can write down the density function of $r_{1,k-1} = (x_n - x_{n-1})/(x_n - x_k)$ as

$$(16) \qquad (n - k - 1)(1 - r_{1,k-1})^{n-k-2},$$

where $k = 0, 1, \cdots, n - 2$.

$n = 4$, *Normal population.* When we assume the normal distribution for our $f(x)$ and consider $k = 2$, the first sample size of interest is $n = 4$, here $r_{11} = (x_4 - x_3)/(x_4 - x_2)$. The density function may be obtained for this ratio by the procedures used above for r_{10}. The helpful substitution here is $r_{11} = (\sqrt{2}/2 + \sqrt{w^2 - 1})^{-1/2}$. The resulting expression is

$$(17) \qquad g(r_{11}) = \frac{3\sqrt{3}}{\pi(r^2 - r + 1)} \left[1 + \frac{r - 2}{\sqrt{3}(4 - 4r + 3r^2)^{1/2}} \right]$$

and the cdf is

$$(18) \qquad \frac{6}{\pi} \left[\arctan \frac{1}{\sqrt{3}} (2R - 1) + \arctan \frac{1}{R} (4 - 4R + 3R^2)^{1/2} \right] - 2.$$

If we now set this function equal to $1 - \alpha$, we may solve for the various percentage values for this distribution.

$n = 5$, *Normal population.* The distribution of the similar ratio for samples of size five, $r_{11} = (x_5 - x_4)/(x_5 - x_2)$ is integrable into an expression similar to the distribution of r_{10} for $n = 5$. The percentage values for the distribution of r_{11} for $n = 4, \cdots, 30$ are in Table II. The distribution of r_{11} for samples of size 5 is

$$\alpha \left[\frac{\beta}{\sqrt{3}} \left(\tan^{-1} \frac{\delta}{\sqrt{5}} - 2 \tan^{-1} \frac{\beta}{\sqrt{5}} \right) - \frac{\pi\gamma}{6} (\beta + \gamma) \tan^{-1} \frac{\delta'}{\sqrt{5}} \right],$$

where the symbols in this expression and those to follow are

$$\alpha = \frac{15\sqrt{3}}{\pi^3(1 - r + r^2)}, \qquad \beta = (2 - r)/q_1, \qquad \delta = (3r - 2)/q_1$$

$$q_1 = \sqrt{4 - 4r + 3r^2}, \qquad \beta' = (2 + r)/q_1, \qquad \delta' = (3 - 2r)/q_2,$$

$$q_2 = \sqrt{3 - 4r + 4r^2}, \qquad \gamma = (1 - 2r)/q_2, \qquad \eta = (1 + r)/q_3,$$

$$q_3 = \sqrt{3 - 2r + 3r^2}, \qquad \gamma' = (1 + 2r)/q_2, \qquad \eta' = (3 - r)/q_3,$$

$$\eta'' = (3r - 1)/q_3.$$

The percentage values of the distribution of the ratio $r_{12} = (x_n - x_1)/(x_n - x_3)$ are in Table III. The general expression for the cdf is

$$\int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{n!}{2(n-4)!} \left(\int_{-\infty}^{x-v} f(t) \, dt \right)^2 \left(\int_{x-v}^{x-R_{12}v} f(s) \, ds \right)^{n-4} f(x - v)f(x) \, dv \, dx.$$

The smallest sample size for which this ratio will have meaning is $n = 5$. The density function for $n = 5$ is

$$\frac{\alpha}{2} \left[\frac{\pi}{2} + \tan^{-1} \frac{1}{\sqrt{15}} + \frac{2\beta}{\sqrt{3}} \tan^{-1} \frac{\beta}{\sqrt{5}} - \frac{\pi\beta}{\sqrt{3}} \right].$$

Percentage values have been computed in a similar manner for $r_{20} = (x_n - x_{n-2})/(x_n - x_1)$, $r_{21} = (x_n - x_{n-2})/(x_n - x_2)$, $r_{22} = (x_n - x_{n-2})/(x_n - x_3)$

and are presented in Tables IV, V, and VI. Here again analytic expressions can be obtained for the distribution of a particular ratio for small values of n.

We have the distribution of r_{20} for $n = 4$ since for this sample size $r_{20} + r_{10} = 1$ if we consider $r_{10} = \dfrac{x_2 - x_1}{x_n - x_1}$.

For $n = 5$ the density function of r_{20} is

$$\alpha\left[\frac{\beta}{\sqrt{3}}\left(\tan^{-1}\frac{\delta}{\sqrt{5}} + \tan^{-1}\frac{\beta'}{\sqrt{5}}\right)\right.$$
$$+ \frac{\gamma}{\sqrt{3}}\left(\tan^{-1}\frac{\gamma'}{\sqrt{5}} - 2\tan^{-1}\frac{\gamma}{\sqrt{5}} - \tan^{-1}\frac{\delta'}{\sqrt{5}}\right)$$
$$\left. + \frac{\eta}{\sqrt{3}}\left(\tan^{-1}\frac{\eta'}{\sqrt{5}} - \tan^{-1}\frac{\eta''}{\sqrt{5}}\right)\right].$$

For $n = 5$ the density function of r_{21} is

$$\alpha\left[\frac{-\beta}{\sqrt{3}}\left(\tan^{-1}\frac{\delta}{\sqrt{5}} + \tan^{-1}\frac{\beta'}{\sqrt{5}}\right) - \frac{\gamma}{\sqrt{3}}\left(\frac{\pi}{2} - \tan^{-1}\frac{\delta'}{\sqrt{5}}\right)\right.$$
$$\left. + \frac{\eta}{\sqrt{3}}\left(\frac{\pi}{2} - \tan^{-1}\frac{\eta'}{\sqrt{5}}\right)\right].$$

The distribution for the ratio $r_{j,i-1} = (x_n - x_j)/(x_n - x_i)$ is

$$\int_{-\infty}^{\infty}\int_{0}^{\infty}\frac{n!}{(i-1)!(n-j-i-1)!(j-1)!}\left(\int_{-\infty}^{x-v}f(t)\,dt\right)^{i-1}f(x-v)$$
$$\cdot\left(\int_{x-v}^{x-rv}f(t)\,dt\right)^{n-j-i-1}f(x-rv)f(x)\left(\int_{x-rv}^{x}f(t)\,dt\right)^{j-1}dv\,dx.$$

5. Final remarks.

5.1. *Accuracy of tables.* The goal with respect to accuracy was to obtain three places of accuracy in the percentage values. It is believed that the values in Tables I, II, III are in error by not more than one or two in the third place, while the values in Tables IV, V, and VI are believed to be accurate to within three or four units in the third place.

5.2. *Investigation of the performance of the ratios.* It is important to know something about the performance of these ratios for various purposes. Reference is made to another paper [1] evaluating the performance of these criteria as well as a number of others.

REFERENCE

[1] W. J. DIXON, "Analysis of extreme values," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 488–506.

TABLE I

$$Pr(r_{10} > R) = \alpha$$

n	α	.005	.01	.02	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	α	n
3		.994	.988	.976	.941	.886	.781	.684	.591	.500	.409	.316	.219	.114	.059		3
4		.926	.889	.846	.765	.679	.560	.471	.394	.324	.257	.193	.130	.065	.033		4
5		.821	.780	.729	.642	.557	.451	.373	.308	.250	.196	.146	.097	.048	.023		5
6		.740	.698	.644	.560	.482	.386	.318	.261	.210	.164	.121	.079	.038	.018		6
7		.680	.637	.586	.507	.434	.344	.281	.230	.184	.143	.105	.068	.032	.016		7
8		.634	.590	.543	.468	.399	.314	.255	.208	.166	.128	.094	.060	.029	.014		8
9		.598	.555	.510	.437	.370	.290	.234	.191	.152	.118	.086	.055	.026	.013		9
10		.568	.527	.483	.412	.349	.273	.219	.178	.142	.110	.080	.051	.025	.012		10
11		.542	.502	.460	.392	.332	.259	.208	.168	.133	.103	.074	.048	.023	.011		11
12		.522	.482	.441	.376	.318	.247	.197	.160	.126	.097	.070	.045	.022	.011		12
13		.503	.465	.425	.361	.305	.237	.188	.153	.120	.092	.067	.043	.021	.010		13
14		.488	.450	.411	.349	.294	.228	.181	.147	.115	.088	.064	.041	.020	.010		14
15		.475	.438	.399	.338	.285	.220	.175	.141	.111	.085	.062	.040	.019	.010		15
16		.463	.426	.388	.329	.277	.213	.169	.136	.107	.082	.060	.039	.019	.009		16
17		.452	.416	.379	.320	.269	.207	.165	.132	.104	.080	.058	.038	.018	.009		17
18		.442	.407	.370	.313	.263	.202	.160	.128	.101	.078	.056	.036	.018	.009		18
19		.433	.398	.363	.306	.258	.197	.157	.125	.098	.076	.055	.036	.017	.008		19
20		.425	.391	.356	.300	.252	.193	.153	.122	.096	.074	.053	.035	.017	.008		20
21		.418	.384	.350	.295	.247	.189	.150	.119	.094	.072	.052	.034	.016	.008		21
22		.411	.378	.344	.290	.242	.185	.147	.117	.092	.071	.051	.033	.016	.008		22
23		.404	.372	.338	.285	.238	.182	.144	.115	.090	.069	.050	.033	.016	.008		23
24		.399	.367	.333	.281	.234	.179	.142	.113	.089	.068	.049	.032	.016	.008		24
25		.393	.362	.329	.277	.230	.176	.139	.111	.088	.067	.048	.032	.015	.008		25
26		.388	.357	.324	.273	.227	.173	.137	.109	.086	.066	.047	.031	.015	.007		26
27		.384	.353	.320	.269	.224	.171	.135	.108	.085	.065	.047	.031	.015	.007		27
28		.380	.349	.316	.266	.220	.168	.133	.106	.084	.064	.046	.030	.015	.007		28
29		.376	.345	.312	.263	.218	.166	.131	.105	.083	.063	.046	.030	.014	.007		29
30		.372	.341	.309	.260	.215	.164	.130	.103	.082	.062	.045	.029	.014	.007		30

TABLE II

$$Pr(r_{11} > R) = \alpha$$

n \ α	.005	.01	.02	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	α / n
4	.995	.991	.981	.955	.910	.822	.737	.648	.554	.459	.362	.250	.131	.069	4
5	.937	.916	.876	.807	.728	.615	.524	.444	.369	.296	.224	.151	.078	.039	5
6	.839	.805	.763	.689	.609	.502	.420	.350	.288	.227	.169	.113	.056	.028	6
7	.782	.740	.689	.610	.530	.432	.359	.298	.241	.189	.140	.093	.045	.022	7
8	.725	.683	.631	.554	.479	.385	.318	.260	.210	.164	.121	.079	.037	.019	8
9	.677	.635	.587	.512	.441	.352	.288	.236	.189	.148	.107	.070	.033	.016	9
10	.639	.597	.551	.477	.409	.325	.265	.216	.173	.134	.098	.063	.030	.014	10
11	.606	.566	.521	.450	.385	.305	.248	.202	.161	.124	.090	.058	.028	.013	11
12	.580	.541	.498	.428	.367	.289	.234	.190	.150	.116	.084	.055	.026	.012	12
13	.558	.520	.477	.410	.350	.275	.222	.180	.142	.109	.079	.052	.025	.012	13
14	.539	.502	.460	.395	.336	.264	.212	.171	.135	.104	.075	.049	.024	.011	14
15	.522	.486	.445	.381	.323	.253	.203	.164	.129	.099	.072	.047	.023	.011	15
16	.508	.472	.432	.369	.313	.244	.196	.158	.124	.095	.069	.045	.022	.011	16
17	.495	.460	.420	.359	.303	.236	.190	.152	.119	.092	.067	.044	.021	.010	17
18	.484	.449	.410	.349	.295	.229	.184	.148	.116	.089	.065	.042	.020	.010	18
19	.473	.439	.400	.341	.288	.223	.179	.143	.112	.087	.063	.041	.020	.010	19
20	.464	.430	.392	.334	.282	.218	.174	.139	.110	.084	.061	.040	.019	.010	20
21	.455	.421	.384	.327	.276	.213	.170	.136	.107	.082	.059	.039	.019	.009	21
22	.446	.414	.377	.320	.270	.208	.166	.132	.104	.081	.058	.038	.018	.009	22
23	.439	.407	.371	.314	.265	.204	.163	.130	.102	.079	.056	.037	.018	.009	23
24	.432	.400	.365	.309	.260	.200	.160	.127	.100	.077	.055	.036	.018	.009	24
25	.426	.394	.359	.304	.255	.197	.156	.124	.098	.076	.054	.036	.017	.009	25
26	.420	.389	.354	.299	.250	.193	.154	.122	.096	.074	.053	.035	.017	.008	26
27	.414	.383	.349	.295	.246	.190	.151	.120	.095	.073	.052	.034	.017	.008	27
28	.409	.378	.344	.291	.243	.188	.149	.118	.093	.072	.051	.034	.016	.008	28
29	.404	.374	.340	.287	.239	.185	.146	.116	.092	.070	.051	.033	.016	.008	29
30	.399	.369	.336	.283	.236	.182	.144	.115	.090	.069	.050	.032	.016	.008	30

TABLE III

$$Pr(r_{12} > R) = \alpha$$

n	α	.005	.01	.02	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	α	n
5		.996	.992	.984	.960	.919	.838	.755	.669	.579	.483	.381	.268	.143	.074		5
6		.951	.925	.891	.824	.745	.635	.545	.465	.390	.316	.240	.165	.088	.049		6
7		.875	.836	.791	.712	.636	.528	.445	.374	.307	.245	.183	.123	.064	.031		7
8		.797	.760	.708	.632	.557	.456	.382	.317	.258	.203	.152	.101	.056	.025		8
9		.739	.701	.656	.580	.504	.409	.339	.270	.227	.177	.130	.086	.044	.021		9
10		.694	.655	.610	.537	.454	.373	.308	.258	.204	.158	.116	.075	.038	.019		10
11		.658	.619	.575	.502	.431	.345	.283	.232	.187	.145	.106	.069	.035	.917		11
12		.629	.590	.546	.473	.406	.324	.265	.217	.174	.135	.098	.063	.032	.016		12
13		.612	.554	.521	.451	.387	.307	.250	.204	.163	.126	.092	.059	.030	.015		13
14		.580	.542	.501	.432	.369	.292	.237	.193	.153	.118	.086	.055	.028	.014		14
15		.560	.523	.482	.416	.354	.280	.226	.184	.146	.112	.082	.053	.026	.013		15
16		.544	.508	.467	.401	.341	.269	.217	.177	.139	.107	.078	.050	.025	.013		16
17		.529	.493	.453	.388	.330	.259	.209	.170	.134	.103	.075	.048	.024	.012		17
18		.516	.480	.440	.377	.320	.251	.202	.163	.129	.099	.072	.047	.023	.012		18
19		.504	.469	.429	.367	.311	.243	.196	.157	.125	.096	.069	.045	.022	.011		19
20		.493	.458	.419	.358	.303	.237	.191	.153	.121	.093	.067	.044	.022	.011		20
21		.483	.449	.410	.349	.296	.231	.186	.148	.118	.090	.065	.042	.021	.010		21
22		.474	.440	.402	.342	.290	.225	.181	.145	.114	.088	.063	.041	.020	.010		22
23		.465	.432	.394	.336	.284	.220	.176	.141	.112	.086	.062	.040	.020	.010		23
24		.457	.423	.387	.330	.278	.216	.173	.138	.109	.084	.060	.039	.019	.010		24
25		.450	.417	.381	.324	.273	.212	.169	.135	.107	.082	.059	.038	.019	.009		25
26		.443	.411	.375	.319	.268	.208	.166	.132	.105	.080	.058	.037	.019	.009		26
27		.437	.405	.370	.314	.263	.204	.163	.130	.103	.079	.057	.037	.018	.009		27
28		.431	.399	.365	.309	.259	.201	.160	.128	.101	.077	.056	.036	.018	.009		28
29		.426	.394	.360	.305	.255	.197	.157	.126	.099	.076	.055	.035	.017	.009		29
30		.420	.389	.355	.301	.251	.194	.154	.124	.098	.075	.054	.035	.017	.009		30

TABLE IV

$$Pr(r_{20} > R) = \alpha$$

n \ α	.005	.01	.02	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	α \ n
4	.996	.992	.987	.967	.935	.871	.807	.743	.676	.606	.529	.440	.321	.235	4
5	.950	.929	.901	.845	.782	.694	.623	.560	.500	.440	.377	.306	.218	.155	5
6	.865	.836	.800	.736	.670	.585	.520	.463	.411	.358	.305	.245	.172	.126	6
7	.814	.778	.732	.661	.596	.516	.454	.402	.355	.306	.261	.208	.144	.099	7
8	.746	.710	.670	.607	.545	.468	.410	.361	.317	.274	.230	.184	.125	.085	8
9	.700	.667	.627	.565	.505	.432	.378	.331	.288	.250	.208	.166	.114	.077	9
10	.664	.632	.592	.531	.474	.404	.354	.307	.268	.231	.192	.153	.104	.070	10
11	.627	.603	.564	.504	.449	.381	.334	.290	.253	.217	.181	.143	.097	.065	11
12	.612	.579	.540	.481	.429	.362	.316	.274	.239	.205	.172	.136	.091	.060	12
13	.590	.557	.520	.461	.411	.345	.301	.261	.227	.195	.164	.129	.086	.057	13
14	.571	.538	.502	.445	.395	.332	.288	.250	.217	.187	.157	.123	.082	.054	14
15	.554	.522	.486	.430	.382	.320	.277	.241	.209	.179	.150	.118	.079	.052	15
16	.539	.508	.472	.418	.370	.310	.268	.233	.202	.173	.144	.113	.076	.050	16
17	.526	.495	.460	.406	.359	.301	.260	.226	.195	.167	.139	.109	.074	.049	17
18	.514	.484	.449	.397	.350	.293	.252	.219	.189	.162	.134	.105	.071	.048	18
19	.503	.473	.439	.379	.341	.286	.246	.213	.184	.157	.130	.101	.069	.047	19
20	.494	.464	.430	.372	.333	.279	.240	.208	.179	.152	.126	.098	.067	.046	20
21	.485	.455	.422	.365	.326	.273	.235	.203	.175	.148	.123	.096	.065	.045	21
22	.477	.447	.414	.358	.320	.267	.230	.199	.171	.145	.120	.094	.064	.044	22
23	.469	.440	.407	.352	.314	.262	.225	.195	.167	.142	.117	.092	.062	.043	23
24	.462	.434	.401	.347	.309	.258	.221	.192	.164	.139	.114	.090	.061	.042	24
25	.456	.428	.395	.343	.304	.254	.217	.189	.161	.136	.112	.089	.060	.041	25
26	.450	.422	.390	.338	.300	.250	.214	.186	.158	.134	.110	.087	.059	.041	26
27	.444	.417	.385	.334	.296	.246	.211	.183	.156	.132	.109	.086	.058	.040	27
28	.439	.412	.381	.330	.292	.243	.208	.180	.154	.130	.107	.085	.058	.040	28
29	.434	.407	.376	.326	.288	.239	.205	.177	.151	.128	.106	.083	.057	.039	29
30	.428	.402	.372	.322	.285	.236	.202	.175	.149	.126	.104	.082	.056	.039	30

TABLE V

$$Pr(r_n > R) = \alpha$$

n \ α	.005	.01	.02	.05	.10	.20	.30	.40	.50	60	.70	.80	.90	.95	α \ n
5	.998	.995	.990	.976	.952	.902	.850	.795	.735	.669	.594	.501	.374	.273	5
6	.970	.951	.924	.872	.821	.745	.680	.621	.563	.504	.439	.364	.268	.195	6
7	.919	.885	.842	.780	.725	.637	.575	.517	.462	.408	.350	.285	.198	.138	7
8	.868	.829	.780	.710	.650	.570	.509	.454	.402	.352	.298	.240	.166	.117	8
9	.816	.776	.725	.657	.594	.516	.458	.407	.360	.313	.265	.212	.146	.103	9
10	.760	.726	.678	.612	.551	.474	.420	.374	.329	.286	.240	.189	.130	.089	10
11	.713	.679	.638	.576	.517	.442	.391	.348	.305	.265	.221	.173	.118	.080	11
12	.675	.642	.605	.546	.490	.419	.370	.326	.285	.247	.206	.161	.110	.074	12
13	.649	.615	.578	.521	.467	.399	.351	.308	.269	.232	.194	.152	.104	.070	13
14	.627	.593	.556	.501	.448	.381	.334	.293	.256	.219	.184	.144	.099	.066	14
15	.607	.574	.537	.483	.431	.366	.319	.280	.245	.208	.175	.138	.094	.062	15
16	.589	.557	.521	.467	.416	.353	.307	.269	.235	.199	.167	.132	.090	.059	16
17	.573	.542	.507	.453	.403	.341	.296	.259	.225	.192	.161	.127	.686	.057	17
18	.559	.529	.494	.4,0	.391	.331	.287	.250	.218	.186	.155	.122	.082	.054	18
19	.547	.517	.482	.428	.380	.322	.279	.243	.211	.180	.150	.117	.078	.052	19
20	.536	.506	.472	.419	.371	.314	.271	.236	.205	.174	.145	.113	.075	.050	20
21	.526	.496	.462	.410	.363	.306	.264	.229	.199	.170	.141	.110	.073	.049	21
22	.517	.487	.453	.402	.356	.299	.258	.223	.194	.165	.137	.107	.071	.048	22
23	.509	.479	.445	.395	.349	.293	.252	.218	.189	.161	.133	.105	.069	.046	23
24	.501	.471	.438	.388	.343	.287	.247	.214	.185	.158	.130	.103	.068	.045	24
25	.493	.464	.431	.382	.337	.282	.242	.210	.181	.154	.127	.100	.067	.043	25
26	.486	.457	.424	.376	.331	.277	.238	.206	.178	.151	.125	.098	.066	.042	26
27	.479	.450	.418	.370	.325	.273	.234	.203	.175	.149	.123	.096	.064	.041	27
28	.472	.444	.412	.365	.320	.269	.230	.200	.172	.146	.121	.094	.063	.041	28
29	.466	.438	.406	.360	.316	.265	.227	.197	.170	.144	.119	.092	.062	.040	29
30	.460	.433	.401	.355	.312	.261	.224	.194	.167	.142	.117	.091	.061	.040	30

TABLE VI

$$Pr(r_{11} > R) = \alpha$$

n \ α	.005	.01	.02	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	α \ n
6	.998	.995	.992	.983	.965	.930	.880	.830	.780	.720	.640	.540	.410	.300	6
7	.970	.945	.919	.881	.850	.780	.730	.670	.610	.540	.470	.390	.270	.200	7
8	.922	.890	.857	.803	.745	.664	.602	.546	.490	.434	.375	.309	.218	.156	8
9	.873	.840	.800	.737	.676	.592	.530	.478	.425	.373	.320	.261	.186	.128	9
10	.826	.791	.749	.682	.620	.543	.483	.433	.384	.335	.285	.231	.150	.111	10
11	.781	.745	.703	.637	.578	.503	.446	.397	.351	.305	.258	.208	.142	.099	11
12	.740	.704	.661	.600	.543	.470	.416	.370	.325	.282	.238	.190	.130	.090	12
13	.705	.670	.628	.570	.515	.443	.391	.347	.304	.263	.222	.177	.122	.084	13
14	.674	.641	.602	.546	.492	.421	.370	.328	.287	.247	.208	.166	.115	.079	14
15	.647	.616	.579	.525	.472	.402	.353	.312	.273	.234	.196	.156	.109	.075	15
16	.624	.595	.559	.507	.454	.386	.338	.298	.261	.223	.186	.148	.104	.071	16
17	.605	.577	.542	.490	.438	.373	.325	.286	.250	.214	.178	.142	.099	.067	17
18	.589	.561	.527	.475	.424	.361	.314	.276	.241	.206	.171	.135	.094	.063	18
19	.575	.547	.514	.462	.412	.350	.304	.268	.233	.199	.165	.130	.090	.060	19
20	.562	.535	.502	.450	.401	.340	.295	.260	.226	.193	.160	.125	.086	.057	20
21	.551	.524	.491	.440	.391	.331	.287	.252	.220	.187	.155	.120	.082	.054	21
22	.541	.514	.481	.430	.382	.323	.280	.245	.213	.182	.150	.116	.078	.051	22
23	.532	.505	.472	.421	.374	.316	.274	.239	.207	.177	.146	.113	.075	.049	23
24	.524	.497	.484	.413	.367	.310	.268	.232	.201	.172	.142	.111	.074	.047	24
25	.516	.489	.457	.406	.360	.304	.262	.227	.196	.168	.138	.108	.073	.045	25
26	.508	.486	.450	.399	.354	.298	.257	.222	.192	.164	.135	.106	.072	.044	26
27	.501	.475	.443	.393	.348	.292	.252	.218	.189	.161	.132	.104	.071	.043	27
28	.495	.469	.437	.387	.342	.287	.247	.215	.186	.158	.130	.102	.069	.042	28
29	.489	.463	.431	.381	.337	.282	.243	.211	.183	.155	.128	.100	.068	.041	29
30	.483	.457	.425	.376	.332	.278	.239	.208	.180	.153	.126	.098	.067	.041	30

ON INFORMATION AND SUFFICIENCY

By S. Kullback and R. A. Leibler

The George Washington University and Washington, D. C.

1. Introduction. This note generalizes to the abstract case Shannon's definition of information [15], [16]. Wiener's information (p. 75 of [18]) is essentially the same as Shannon's although their motivation was different (cf. footnote 1, p. 95 of [16]) and Shannon apparently has investigated the concept more completely. R. A. Fisher's definition of information (intrinsic accuracy) is well known (p. 709 of [6]). However, his concept is quite different from that of Shannon and Wiener, and hence ours, although the two are not unrelated as is shown in paragraph 2.

R. A. Fisher, in his original introduction of the *criterion of sufficiency*, required "that the statistic chosen should summarize the whole of the relevant information supplied by the sample," (p. 316 of [5]). Halmos and Savage in a recent paper, one of the main results of which is a generalization of the well known Fisher-Neyman theorem on sufficient statistics to the abstract case, conclude, "We think that confusion has from time to time been thrown on the subject by ... , and (c) the assumption that a sufficient statistic contains all the information in only the technical sense of 'information' as measured by variance," (p. 241 of [8]). It is shown in this note that the information in a sample as defined herein, that is, in the Shannon-Wiener sense cannot be increased by any statistical operations and is invariant (not decreased) if and only if sufficient statistics are employed. For a similar property of Fisher's information see p. 717 of [6], Doob [19].

We are also concerned with the statistical problem of discrimination ([3], [17]), by considering a measure of the "distance" or "divergence" between statistical populations ([1], [2], [13]) in terms of our measure of information. For the statistician two populations differ more or less according as to how difficult it is to discriminate between them with the best test [14]. The particular measure of divergence we use has been considered by Jeffreys ([10], [11]) in another connection. He is primarily concerned with its use in providing an invariant density of *a priori* probability. A special case of this divergence is Mahalanobis' generalized distance [13].

We shall use the notation of Halmos and Savage [8] and that of [7].

2. Information. Assume given the probability spaces $(X, \mathbf{S}, \mu_i)$, $i = 1, 2$, such that $\mu_1 \equiv \mu_2$[1] (cf. p. 228 of [8]) and let λ be a probability measure such that $\lambda \equiv \{\mu_1, \mu_2\}$ (e.g., λ may be μ_1, or μ_2 or $\frac{1}{2}(\mu_1 + \mu_2)$, etc.). By the Radon-Nikodym theorem [7] there exist $f_i(x)$, $i = 1, 2$, unique up to sets of measure zero in λ,

---

[1] If $\mu_1(E) \neq 0$, $\mu_2(E) = 0$ or $\mu_1(E) = 0$, $\mu_2(E) \neq 0$ for $E \cdot \mathbf{S}$ then we can discriminate perfectly between the populations. The assumption $\mu_1 \equiv \mu_2$ that is, that μ_1 and μ_2 are absolutely continuous with respect to each other is made to avoid this situation.

measurable λ with $0 < f_i(x) < \infty$ $[\lambda]$, $i = 1, 2$, such that

$$(2.1) \qquad \mu_i(E) = \int_E f_i(x) \, d\lambda(x), \qquad\qquad i = 1, 2,$$

for all $E \epsilon$ S. If H_i, $i = 1, 2$, is the hypothesis that x was selected from the population whose probability measure is μ_i, $i = 1, 2$ then we define

$$(2.2) \qquad \log \frac{f_1(x)}{f_2(x)}$$

as *the information[2] in x for discrimination between H_1 and H_2*. The mean information for discrimination between H_1 and H_2 per observation from $E \epsilon$ S for μ_1 is given by (cf. pp. 18, 19 of [16]; p. 76 of [18])

$$I_{1:2}(E) = \frac{1}{\mu_1(E)} \int_E d\mu_1(x) \log \frac{f_1(x)}{f_2(x)} = \frac{1}{\mu_1(E)} \int_E f_1(x) \log \frac{f_1(x)}{f_2(x)} \, d\lambda(x)$$

$$(2.3) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for} \quad \mu_1(E) > 0,$$

$$= 0 \qquad\qquad\qquad\qquad \text{for} \quad \mu_1(E) = 0.$$

It should be noted that $I_{1:2}(E)$ in (2.3) is well defined in that the integral in its definition always exists even though it may be $+\infty$, since the measures are finite measures.[3] It is shown in Lemma 3.2 that

$$I_{1:2}(E) \geqq \log \mu_1(E)/\mu_2(E) \quad \text{for} \quad \mu_1(E) > 0.$$

We shall denote by $I(1:2)$ the mean information for discrimination between H_1 and H_2 per observation from μ_1; i.e.,[4]

$$(2.4) \qquad I(1:2) = I_{1:2}(\mathfrak{X}) = \int d\mu_1(x) \log \frac{f_1(x)}{f_2(x)}$$

$$= \int f_1(x) \log \frac{f_1(x)}{f_2(x)} \, d\lambda(x).$$

---

[2] It follows from Bayes' Theorem [12] that

$$\log \frac{f_1(x)}{f_2(x)} = \log \frac{P(H_1 \mid x)}{P(H_2 \mid x)} - \log \frac{\alpha_1}{\alpha_2} \, [\lambda]$$

where α_i, $i = 1, 2$, are the *a priori* probabilities and $P(H_i \mid x)$, $i = 1, 2$, the *a posteriori* probabilities of H_i, $i = 1, 2$, respectively.

[3] We are indebted to a referee for this remark as well as for the following example which shows that the assumptions at the beginning of this paragraph do not imply finiteness of information. Take $E = (0, 1)$, $\mu_1 =$ Lebesgue measure, $f_2(x)/f_1(x) = ke^{-1/x}$, where $k^{-1} = \int_0^1 e^{-1/t} \, dt$. It is easily verified that $I(1:2)$ is infinite (cf. also p. 137 [9]).

[4] We shall omit the region of integration when it is the entire space.

Set

$$J_{12}(E) = I_{112}(E) + I_{221}(E)$$

$$(2.5) \qquad = \frac{1}{\mu_1(E)} \int_E d\mu_1(x) \log \frac{f_1(x)}{f_2(x)} + \frac{1}{\mu_2(E)} \int_E d\mu_2(x) \log \frac{f_2(x)}{f_1(x)}$$

$$= \int_E \left(\frac{f_1(x)}{\mu_1(E)} - \frac{f_2(x)}{\mu_2(E)} \right) \log \frac{f_1(x)}{f_2(x)} \, d\lambda(x).$$

We denote by $J(1, 2)$ the "divergence" between μ_1 and μ_2 (cf. p. 158 of [11]) so that[4]

$$(2.6) \qquad J(1, 2) = J_{12}(X) = \int (f_1(x) - f_2(x)) \log \frac{f_1(x)}{f_2(x)} \, d\lambda(x).$$

Shannon ([15], [16]) defined information on a finite discrete space and we note that $I_{1:2}(E)$ defined in (2.3) is precisely the generalization of that information which is obtained when one replaces the finite space by $S \cap E$, the measure of equidistribution by $\mu_2/\mu_1(E)$ and the measure whose information is being defined by $\mu_1/\mu_1(E)$. Just as Shannon observed that certain theorems were carried over to the Lebesgue case, we shall see here that they maybe formally carried over to the general case.[5]

For the parametric case in which $f_1(x) = f(x, \theta)$ and $f_2(x) = f(x, \theta + \Delta\theta)$, where θ and $\theta + \Delta\theta$ are neighboring points in the k-dimensional parameter space, with suitable assumptions on the density function (e.g., see p. 774 of [4]), to within second order terms it is found that

$$(2.7) \qquad\qquad I(\theta; \theta + \Delta\theta) = \tfrac{1}{2} \Sigma g_{\alpha\beta} \Delta\theta_\alpha \Delta\theta_\beta, \qquad \alpha, \beta = 1, \cdots, k,$$

$$(2.8) \qquad\qquad J(\theta, \theta + \Delta\theta) = \Sigma g_{\alpha\beta} \Delta\theta_\alpha \Delta\theta_\beta, \qquad \alpha, \beta = 1, \cdots, k,$$

where

$$(2.9) \qquad\qquad g_{\alpha\beta} = \int f \left(\frac{1}{f} \frac{\partial f}{\partial \theta_\alpha} \right) \left(\frac{1}{f} \frac{\partial f}{\partial \theta_\beta} \right) d\lambda$$

are the elements of Fisher's information matrix (cf. par. 3.9 of [11]).

When μ_1 and μ_2 are multivariate normal populations with a common matrix of variances and covariances then

$$(2.10) \qquad\qquad J(1, 2) = \Sigma \delta_\alpha \delta_\beta \sigma^{\alpha\beta}, \qquad\qquad \alpha, \beta = 1, \cdots, k,$$

where δ_α, $\alpha = 1, \cdots, k$, are the differences of the respective population means and $\sigma^{\alpha\beta}$, $\alpha, \beta = 1, \cdots, k$, are the elements of the inverse of the common matrix

---

[5] We are indebted to a referee for the comments with respect to Shannon's definition as well as for the comment that this should be of interest to anyone who has puzzled over Wiener's statement that his definition of "information" can be used to replace Fisher's definition in the technique of statistics (p. 76 of [18]).

of variances and covariances; i.e., $J(1, 2)$ in (2.10) is k times Mahalanobis' generalized distance [13].

3. Some properties of information.

LEMMA 3.1. $I(1:2)$ is almost positive definite; i.e., $I(1:2) \geqq 0$ with equality if and only if $f_1(x) = f_2(x)$ [λ].

PROOF.[6] Let $g(x) = f_1(x)/f_2(x)$. Then

$$
\begin{aligned}
(3.1) \quad I(1:2) &= \int f_2(x)g(x) \log g(x) \, d\lambda(x) \\
&= \int g(x) \log g(x) \, d\mu_2(x).
\end{aligned}
$$

If we write $\varphi(t) = t \log t$, then since $0 < g(x) < \infty$ [λ] and

$$
(3.2) \qquad \int g(x) \, d\mu_2(x) = \int f_1(x) \, d\lambda(x) = 1,
$$

we may write

$$
(3.3) \qquad \varphi(g(x)) = \varphi(1) + [g(x) - 1]\varphi'(1) + \tfrac{1}{2}[g(x) - 1]^2\varphi''(h(x))[\lambda],
$$

where $h(x)$ lies between $g(x)$ and 1 so that $0 < h(x) < \infty$ [λ]. Therefore

$$
(3.4) \qquad \int \varphi(g(x)) \, d\mu_2(x) = \frac{1}{2} \int [g(x) - 1]^2 \varphi''(h(x)) \, d\mu_2(x),
$$

where $\varphi''(t) = \frac{1}{t} > 0$ for $t > 0$. It therefore follows from (3.4) that

$$
(3.5) \qquad \int g(x) \log g(x) \, d\mu_2(x) \geqq 0
$$

with equality if and only if $g(x) = 1$ [λ].

LEMMA 3.2.

$$
I_{1:2}(E) \geqq \log \frac{\mu_1(E)}{\mu_2(E)} \qquad \text{for } \lambda(E) > 0.
$$

PROOF. If $I_{1:2}(E) = \infty$, the result is trivial. For finite $I_{1:2}(E)$ apply Lemma 3.1 to

$$
I_{1:2}(E) - \log \frac{\mu_1(E)}{\mu_2(E)} = \int_E \frac{d\mu_1(x)}{\mu_1(E)} \log \frac{f_1(x)/\mu_1(E)}{f_2(x)/\mu_2(E)}.
$$

THEOREM 3.1. $I(1:2)$ is additive for independent random events[7]; i.e.,

$$
I_{xy}(1:2) = I_x(1:2) + I_y(1:2).
$$

---

[6] This is essentially the proof on p. 151 of [9].

[7] Shannon (p. 21 of [16]) and Wiener (p. 77 of [18]) prove similar results. This is clearly a fundamental property which information must possess, and is one of the *a priori* requirements set down by Shannon in arriving at his definition.

PROOF.

$$I_{xy}(1:2) = \int f_1(x, y) \log \frac{f_1(x, y)}{f_2(x, y)} \, d\lambda(x, y)$$

$$(3.6) \qquad = \iint f_1^{(1)}(x) f_1^{(2)}(y) \log \frac{f_1^{(1)}(x) f_1^{(2)}(y)}{f_2^{(1)}(x) f_2^{(2)}(y)} \, d\lambda_1(x) \, d\lambda_2(y)$$

$$= \int f_1^{(1)}(x) \log \frac{f_1^{(1)}(x)}{f_2^{(1)}(x)} \, d\lambda_1(x) + \int f_1^{(2)}(y) \log \frac{f_1^{(2)}(y)}{f_2^{(2)}(y)} \, d\lambda_2(y)$$

$$= I_x(1:2) + I_y(1:2).$$

4. Transformations and invariance of I(1:2). Consider the measurable transformation T of the probability spaces (X, S, μ_i) *onto* the probability spaces (Y, T, ν_i) and suppose for $G \in T$, $\nu_i(G) = \mu_i(T^{-1}G)$, $i = 1, 2$. Then $\nu_1 \equiv \nu_2 \equiv \gamma$, where $\gamma = \lambda T^{-1}$. We define

$$(4.1) \quad I'_{1:2}(G) = \frac{1}{\nu_1(G)} \int_G d\nu_1(y) \log \frac{g_1(y)}{g_2(y)} = \frac{1}{\nu_1(G)} \int_G g_1(y) \log \frac{g_1(y)}{g_2(y)} \, d\gamma(y),$$

$$(4.2) \qquad J'_{12}(G) = \int_G \left(\frac{d\nu_1(y)}{\nu_1(G)} - \frac{d\nu_2(y)}{\nu_2(G)} \right) \log \frac{g_1(y)}{g_2(y)},$$

where $g_i(y)$ is defined by

$$(4.3) \qquad \nu_i(G) = \int_G g_i(y) \, d\gamma(y), \qquad\qquad i = 1, 2,$$

for all $G \in T$.

THEOREM 4.1. $I(1:2) \geqq I'(1:2)$, *with equality if and only if T is a sufficient statistic.*

PROOF. If $I(1:2) = \infty$ the result is trivial. By Lemma 3 of Halmos and Savage [8]

$$(4.4) \qquad I'(1:2) = \int d\mu_1(x) \log \frac{g_1 T(x)}{g_2 T(x)}.$$

Then

$$I(1:2) - I'(1:2) = \int d\mu_1(x) \left[\log \frac{f_1(x)}{f_2(x)} - \log \frac{g_1 T(x)}{g_2 T(x)} \right]$$

$$(4.5) \qquad\qquad = \int f_1(x) \log \frac{f_1(x) g_2 T(x)}{f_2(x) g_1 T(x)} \, d\lambda(x).$$

If we set $g(x) = \dfrac{f_1(x)g_2\,T(x)}{f_2(x)g_1\,T(x)}$, then

(4.6)
$$I(1{:}2) - I'(1{:}2) = \int \frac{f_2(x)g_1 T(x)}{g_2 T(x)}\, g(x) \log g(x)\, d\lambda(x)$$

$$= \int g(x) \log g(x)\, d\mu_{12}(x),$$

where $\mu_{12}(E) = \displaystyle\int_E \frac{f_2(x)g_1\,T(x)}{g_2\,T(x)}\, d\lambda(x)$ for all $E \,\epsilon\, \mathbf{S}$.

Since

$$\int g(x)\, d\mu_{12}(x) = \int \frac{f_1(x)g_2 T(x)}{f_2(x)g_1 T(x)}\, \frac{f_2(x)g_1 T(x)}{g_2 T(x)}\, d\lambda(x) = 1,$$

the method of Lemma 3.1 leads to the conclusion that $I(1{:}2) - I'(1{:}2) \geqq 0$ with equality if and only if

(4.7)
$$\frac{f_1(x)}{f_2(x)} = \frac{g_1 T(x)}{g_2 T(x)}\ [\lambda].$$

But (4.7) implies that

(4.8)
$$\frac{f_1(x)}{f_2(x)}\ (\varepsilon)\, T^{-1}(\mathbf{T})\ [\lambda],$$

which is by Corollary 2 of Halmos and Savage [8] necessary and sufficient that the statistic T be sufficient for a homogeneous set of measures on $\mathbf{S}$. If T is sufficient then by the same proof[3] as Theorem 1 of Halmos and Savage [8] $f_1(x)$ and $f_2(x)$ are $(\varepsilon)\,T^{-1}(\mathbf{T})[\lambda]$. Then by Lemma 2 of Halmos and Savage [8] and the definition of g_1 and g_2, $f_1(x) = g_1 T(x)\ [\lambda]$, $f_2(x) = g_2 T(x)\ [\lambda]$ and the result in (4.7) follows.

COROLLARY 4.1. $I(1{:}2) = I'(1{:}2)$ *if T is non-singular.*

PROOF. If T is non-singular, $T^{-1}(\mathbf{T})$ is $\mathbf{S}$ and therefore $f_i(x)(\varepsilon)T^{-1}(\mathbf{T})$, $i = 1, 2$. The result then follows from Theorem 4.1.

THEOREM 4.2.[4] $I_{1:2}(T^{-1}G) = I'_{1:2}(G)$ *for all $G\ \epsilon\ \mathbf{T}$ if and only if*

$$I(1{:}2) = I'(1{:}2).$$

PROOF.

(4.9)
$$I'_{1:2}(G) = \int_G \frac{d\nu_1(y)}{\nu_1(G)} \log \frac{g_1(y)}{g_2(y)} = \int \chi_G(y)\, \frac{d\nu_1(y)}{\nu_1(G)} \log \frac{g_1(y)}{g_2(y)}$$

$$= \int \chi_{T^{-1}G}(x)\, \frac{d\mu_1(x)}{\mu_1(T^{-1}G)} \log \frac{g_1 T(x)}{g_2 T(x)}$$

$$= \int_{T^{-1}G} \frac{d\mu_1(x)}{\mu_1(T^{-1}G)} \log \frac{g_1 T(x)}{g_2 T(x)}.$$

Application of the method of Theorem 4.1 completes the proof.

---

[3] Note that the λ in Theorem 1 of [8] is different from the λ here. However, as remarked by a referee, the same proof will suffice.

[4] We are indebted to a referee for calling this to our attention.

5. Properties of $J(1, 2)$. For each of the results in paragraphs 3 and 4 there can be stated an identical one for $J(1, 2)$. This follows from its definition in (2.5) and (2.6). Also it should be noted that $J(1, 2)$ is symmetric with respect to μ_1 and μ_2 and independent of the *a priori* probabilities. Jeffreys (par. 3.9 of [11]) mentioned the symmetry, positive definiteness and additivity, and invariance for non-singular transformations.

6. Application. Two indications of simple application of these concepts may be useful.

(1). Consider the problem of testing an hypothesis presented by Lehmann (p. 2 of [20]). Let the subscript 1 refer to Lehmann's hypothesis H, the subscript 2 refer to any of the alternatives, $F = \{-2, 2\}$, $G = \{0\}$; then

$$I_{1:2}(F) = \frac{1}{\alpha}\left(\frac{\alpha}{2}\log\frac{\alpha}{2pc} + \frac{\alpha}{2}\log\frac{\alpha}{2c(1-p)}\right),$$
(6.1)
$$I_{1:2}(G) = \frac{1}{\alpha} \cdot \alpha \log\frac{1-\alpha}{1-c}.$$

It may be readily verified that $I_{1:2}(G) < I_{1:2}(F)$ and therefore G i.e. $\{0\}$ should be used as the critical region.

(2). Suppose it is necessary to decide whether a sample of n observations has been drawn from the multinomial population $\{p_1, p_2, \cdots, p_k\}$ or $\left\{\frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right\}$. Because of certain limitations the test must be made under the following conditions:

a) Sequential analysis cannot be used.

b) The observations must be grouped into two mutually exclusive categories.

If it is assumed that $p_1 \geqq p_2 \geqq \cdots \geqq p_k$, then the most effective grouping is such that

$$J' = \left(\sum_{i=1}^{r} p_i - \frac{r}{k}\right)\log\frac{\sum_{i=1}^{r} p_i}{r/k} + \left(\sum_{i=r+1}^{k} p_i - \frac{k-r}{k}\right)\log\frac{\sum_{i=r+1}^{k} p_i}{(k-r)/k}$$
(6.2)

is a maximum. The efficiency of the grouped test is measured by

(6.3)
$$J'/J,$$

where

(6.4)
$$J = \sum_{i=1}^{k}\left(p_i - \frac{1}{k}\right)\log\frac{p_i}{1/k}$$

in the sense that n observations of the grouped test will provide as much information as N observations of the ungrouped test where

(6.5)
$$nJ' = NJ.$$

For example if $p_1 = .5$, $p_2 = .3$, $p_3 = .1$, $p_4 = .1$, then using logarithms to base 10, J' for $r = 1, 2, 3, 4$, becomes respectively

(6.6)
$$\qquad\qquad .1193, \qquad .0903, \qquad .0716, \qquad 0.0,$$

and in this case J is 0.1986. The most effective grouping is therefore (p_1), $(p_2 + p_3 + p_4)$ and the grouped case is $\dfrac{.1193}{.1986} = .6007$ times as efficient as the ungrouped test; i.e., there is a loss of 40% because of the grouping.

REFERENCES

[1] A. BHATTACHARYYA, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, Vol. 35 (1943), pp. 99–109.

[2] A. BHATTACHARYYA, "On a measure of divergence between two multinomial populations," *Sankhyā*, Vol. 7 (1946), pp. 401–406.

[3] G. W. BROWN, "Basic principles for construction and application of discriminators," *Jour. Clinical Psych.*, Vol. 6 (1950), pp. 58–61.

[4] J. L. DOOB, "Probability and statistics," *Trans. Am. Math. Soc.*, Vol. 36 (1934), pp. 759–775.

[5] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London, Ser. A*, Vol. 222 (1921), pp. 309–368.

[6] R. A. FISHER, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22 (1925), pp. 700–725.

[7] P. R. HALMOS, *Measure Theory*, D. Van Nostrand, 1950.

[8] P. R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 225–241.

[9] G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, *Inequalities*, Cambridge University Press, 1934.

[10] H. JEFFREYS, "An invariant form for the prior probability in estimation problems," *Proc. Roy. Soc. London, Ser. A.*, Vol. 186 (1946), pp. 453–461.

[11] H. JEFFREYS, *Theory of Probability*, 2nd ed., Oxford, 1948.

[12] A. KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Julius Springer, Berlin, 1933.

[13] P. C. MAHALANOBIS, "On the generalized distance in statistics," *Proc. Nat. Inst. of Sciences of India*, Vol. 2 (1) (1936).

[14] E. MOURIER, "Étude du choix entre deux lois de probabilité," *C. R. Acad. Sci. Paris*, Vol. 223 (1946), pp. 712–714.

[15] C. E. SHANNON, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379–423; pp. 623–656.

[16] C. E. SHANNON AND W. WEAVER, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[17] B. L. WELCH, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218–219.

[18] N. WIENER, *Cybernetics*, John Wiley and Sons, 1948.

[19] J. L. DOOB, "Statistical estimation," *Trans. Am. Math. Soc.*, Vol. 39 (1936), pp. 410–421.

[20] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 1–26.

ON THE FUNDAMENTAL LEMMA OF NEYMAN AND PEARSON[1]

By George B. Dantzig and Abraham Wald[2]

Department of the Air Force and Columbia University

1. Summary and introduction. The following lemma proved by Neyman and Pearson [1] is basic in the theory of testing statistical hypotheses:

LEMMA. *Let* $f_1(x), \cdots, f_{m+1}(x)$ *be* $m + 1$ *Borel measurable functions defined over a finite dimensional Euclidean space* R *such that* $\int_R |f_i(x)| dx < \infty$ ($i = 1, \cdots, m + 1$). *Let, furthermore,* $c_1, \cdots, c_m$ *be* m *given constants and* S *the class of all Borel measurable subsets* S *of* R *for which*

$$(1.1) \qquad \int_S f_i(x) \, dx = c_i \qquad (i = 1, \cdots, m).$$

Let, finally, S_0 *be the subclass of* S *consisting of all members* S_0 *of* S *for which*

$$(1.2) \qquad \int_{S_0} f_{m+1}(x) \, dx \geqq \int_S f_{m+1}(x) \, dx \quad \text{for all } S \text{ in } S.$$

If S *is a member of* S *and if there exist* m *constants* $k_1, \cdots, k_m$ *such that*

$$(1.3) \qquad f_{m+1}(x) \geqq k_1 f_1(x) + \cdots + k_m f_m(x) \qquad \text{when} \quad x \, \epsilon \, S,$$

$$(1.4) \qquad f_{m+1}(x) \leqq k_1 f_1(x) + \cdots + k_m f_m(x) \qquad \text{when} \quad x \, \epsilon \, S,$$

then S *is a member of* S_0.

The above lemma gives merely a sufficient condition for a member S of S to be also a member of S_0. Two important questions were left open by Neyman and Pearson: (1) the question of existence, that is, the question whether S_0 is non-empty whenever S is non-empty; (2) the question of necessity of their sufficient condition (apart from the obvious weakening that (1.3) and (1.4) may be violated on a set of measure zero).

The purpose of the present note is to answer the above two questions. It will be shown in Section 2 that S_0 is not empty whenever S is not empty. In Section 3, a necessary and sufficient condition is given for a member of S to be also a member of S_0. This necessary and sufficient condition coincides with the Neyman-Pearson sufficient condition under a mild restriction.

2. Proof that S_0 is not empty whenever S is not empty. Each function $f_i(x)$ determines a finite measure μ_i given by the equation

$$(2.1) \qquad \mu_i(S) = \int_S f_i(x) \, dx \qquad (i = 1, 2, \cdots, m + 1).$$

Let μ be the vector measure with the components $\mu_1, \cdots, \mu_{m+1}$; i.e., for any measurable set S the value of $\mu(S)$ is the vector $(\mu_1(S), \cdots, \mu_{m+1}(S))$. Thus, for each S the value of $\mu(S)$ can be represented by a point in the $m + 1$-dimensional Euclidean space E. A point $g = (g_1, \cdots, g_{m+1})$ of E is said to belong to the range of the vector measure μ if and only if there exists a measurable subset S of R such that $\mu(S) = g$.

It was proved by Lyapunov [2] (see also [4]) that the range M of μ is a bounded, closed and convex subset of E. Let L be the line in E which is parallel to the $(m + 1)$-th axis and goes through the point $(c_1, c_2, \cdots, c_m, 0)$. Suppose that S is not empty. Then the intersection M^* of L with M is not empty. Because of Lyapunov's theorem, M^* is a finite closed interval (which may reduce to a single point). There exists a subset S of R such that $\mu(S)$ is equal to the upper end point of M^*. Clearly, S is a member of S_0.

3. Necessary and sufficient condition that a member of S be also a member of S_0. Let $\nu(S)$ be the vector measure with the components $\mu_1(S), \cdots \mu_m(S)$. According to the aforementioned theorem of Lyapunov, the range N of ν is a bounded, closed and convex subset of the m-dimensional Euclidean space.

By the dimension of a convex subset Q of a finite dimensional Euclidean space we shall mean the dimension of the smallest dimensional hyperplane that contains Q. A point q of a convex set Q is said to be an interior point of Q if there exists a sphere V with center at q and positive radius such that $V \cap \Pi \subset Q$, where Π is the smallest dimensional hyperplane containing Q. Any point q that is not an interior point of Q will be called a boundary point. We shall now prove the following theorem.

THEOREM 3.1. *If $(c_1, \cdots, c_m)$ is an interior point of N, then a necessary and sufficient condition for a member S of S to be a member of S_0 is that there exist m constants $k_1, \cdots, k_m$ such that (1.3) and (1.4) hold for all x except perhaps on a set of measure zero.*

PROOF. The Neyman-Pearson lemma cited in Section 1 states that our condition is sufficient. Thus, we merely have to prove the necessity of our condition. Assume that $(c_1, \cdots, c_m)$ is an interior point of N. Let c^* be the largest value for which $(c_1, \cdots, c_m, c^*) \epsilon M$, and c^{**} the smallest value for which

$$(c_1, \cdots, c_m, c^{**}) \epsilon M.$$

We shall first consider the case when $c^* = c^{**}$. Let $(\bar{c}_1, \cdots, \bar{c}_m)$ be any other interior point of N. We shall show that there exists exactly one real value $\bar{c}$ such that $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}) \epsilon M$. For suppose that there are two different values $\bar{c}^*$ and $\bar{c}^{**}$ such that both $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}^*)$ and $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}^{**})$ are in M. Since $(c_1, \cdots, c_m)$ and $(\bar{c}_1, \cdots, \bar{c}_m)$ are interior points of N, there exists a point $(c_1', \cdots, c_m')$ in N such that $(c_1, \cdots, c_m)$ lies in the interior of the segment determined by $(c_1', \cdots, c_m')$ and $(\bar{c}_1, \cdots, \bar{c}_m)$. There exists a real value c' such that $(c_1', \cdots, c_m', c') \epsilon M$. Consider the convex set T determined by the 3 points: $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}^*)$, $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}^{**})$ and $(c_1', \cdots, c_m', c')$. Obviously, $T \subset M$. But T contains points $(c_1, \cdots, c_m, h)$ and $(c_1, \cdots, c_m, h')$ with

$h \neq h'$, contrary to our assumption that $c^* = c^{**}$. Thus, for any interior point $(\bar{c}_1, \cdots, \bar{c}_m)$ of N there exists exactly one real value $\bar{c}$ such that $(\bar{c}_1, \cdots, \bar{c}_m, \bar{c}) \epsilon$ M. Since M is closed and convex, this remains true also when $(\bar{c}_1, \cdots, \bar{c}_m)$ is a boundary point of N. Thus, there exists a single valued function $\varphi(g_1, \cdots, g_m)$ such that $g_{m+1} = \varphi(g_1, \cdots, g_m)$ holds for all points $g = (g_1, \cdots, g_m, g_{m+1})$ in M. Since M is convex, φ must be linear; i.e., $\varphi(g_1, \cdots, g_m) = \sum_{i=1}^{m} k_i g_i + k_0$. Since the origin is obviously contained in M, we have $k_0 = 0$. Thus, we have $g_{m+1} = \sum_{i=1}^{m} k_i g_i$ for all points g in M. But then $f_{m+1}(x) = \sum_{i=1}^{m} k_i f_i(x)$ must hold for all x, except perhaps on a set of measure zero. Thus, for any subset S of R, the inequalities (1.3) and (1.4) are fulfilled for all x, except perhaps on a set of measure zero. This completes the proof of our theorem in the case when $c^* = c^{**}$.

We shall now consider the case when $c^{**} < c^*$. Let c be any value between c^{**} and c^*; i.e., $c^{**} < c < c^*$. We shall show that $(c_1, \cdots, c_m, c)$ is an interior point of M. For this purpose, consider a finite set of points $c^i = (c_1^i, \cdots, c_m^i)$ in $N(i = 1, \cdots, n)$ such that $c^1, \cdots, c^n$ are linearly independent, the simplex determined by $c^1, \cdots, c^n$ has the same dimension as N and contains the point $(c_1, \cdots, c_m)$ in its interior. Such points c^i in N obviously exist. There exist real values $h_i(i = 1, \cdots, n)$ such that $(c_1^i, \cdots, c_m^i, h_i) \epsilon M$ $(i = 1, \cdots, n)$. Let T be the smallest convex set containing the points $(c_1^i, \cdots, c_m^i, h_i)$ $(i = 1, \cdots, n)$, $(c_1, \cdots, c_m, c^*)$ and $(c_1, \cdots, c_m, c^{**})$. Clearly, the dimension of T is the same as that of M and $(c_1, \cdots, c_m, c)$ is an interior point of T. Thus, $(c_1, \cdots, c_m, c)$ is an interior point of M. The point $(c_1, \cdots, c_m, c^*)$ is obviously a boundary point of M. Let $g = (g_1, \cdots, g_{m+1})$ be the generic designation of a point in the $m + 1$-dimensional Euclidean space E. Since $(c_1, \cdots, c_m, c^*)$ is a boundary point of M, there exists an m-dimensional hyperplane Π through $(c_1, \cdots, c_m, c^*)$ such that Π contains only boundary points of M and M lies entirely on one side of Π.[3] Let the equation of Π be given by

$$(3.1) \qquad k_{m+1} g_{m+1} - \sum_{i=1}^{m} k_i g_i = k_{m+1} c^* - \sum_{i=1}^{m} k_i c_i.$$

Since Π contains only boundary points of M, and since $(c_1, \cdots, c_m, c)$ is not a boundary point when $c^{**} < c < c^*$, the hyperplane Π cannot be parallel to the $(m + 1)$-th coordinate axis; i.e., $k_{m+1} \neq 0$. We can assume without loss of generality that $k_{m+1} = 1$. Since M lies entirely on one side of Π, and since for $(g_1, \cdots, g_m, g_{m+1}) = (c_1, \cdots, c_m, c^{**})$ the left hand member of (3.1) is smaller than the right hand member, we must have

$$(3.2) \qquad g_{m+1} - \sum_{i=1}^{m} k_i g_i \leq c^* - \sum_{i=1}^{m} k_i c_i$$

for all $g \epsilon M$. Let S be a subset of R such that

---

[3] This follows from well known results on convex bodies. See, for example, [3], p. 6.

$$(3.3) \qquad (\mu_1(S), \cdots, \mu_m(S), \mu_{m+1}(S)) = (c_1, \cdots, c_m, c^*).$$

It can easily be seen that (3.2) and (3.3) can be fulfilled simultaneously only if S satisfies the conditions (1.3) and (1.4) for all x, except perhaps on a set of measure zero. This completes the proof of our theorem.

It remains to investigate the case when $(c_1, \cdots, c_m)$ is a boundary point of N. For this purpose, we shall introduce some definitions and prove some lemmas.

Let $\xi = (\xi_1, \cdots, \xi_m)$ be an m-dimensional vector with real valued components at least one of which is not zero. We shall say that ξ is maximal relative to the point $c = (c_1, \cdots, c_m)$ if

$$(3.4) \qquad \sum_{i=1}^{m} \xi_i g_i \leqq \sum_{i=1}^{m} \xi_i c_i$$

for all points $(g_1, \cdots, g_m)$ in N.

We shall say that a set $\{\xi^i\} (i = 1, 2, \cdots, r; r > 1)$ of vectors is maximal relative to the point $c = (c_1, \cdots, c_m)$ if the set $\{\xi^i\} (i = 1, \cdots, r - 1)$ is maximal relative to c, not all components of ξ^r are zero and

$$(3.5) \qquad \sum_{j=1}^{m} \xi_j^r g_j \leqq \sum_{j=1}^{m} \xi_j^r c_j$$

holds for all points $(g_1, \cdots, g_m)$ of N for which

$$(3.6) \qquad \sum_{j=1}^{m} \xi_j^i g_j = \sum_{j=1}^{m} \xi_j^i c_j \qquad (i = 1, \cdots, r - 1).$$

A set of vectors $\{\xi^i\} (i = 1, \cdots, r)$ is said to be a complete maximal set relative to $c = (c_1, \cdots, c_m)$ if $\{\xi^i\} (i = 1, 2, \cdots, r)$ is maximal relative to c and no vector ξ^{r+1} exists such that ξ^{r+1} is linearly independent of the sequence $(\xi^1, \cdots, \xi^r)$ and $(\xi^1, \cdots, \xi^r, \xi^{r+1})$ is maximal relative to c.

LEMMA 3.1. *If $c = (c_1, \cdots, c_m)$ is a boundary point of N, then there exists a positive integer r and a set $\{\xi^1, \cdots, \xi^r\}$ of vectors that is a complete maximal set relative to c.*

PROOF. Since c is a boundary point of N, there exists an $(m - 1)$-dimensional hyperplane Π through c such that N lies entirely on one side of Π.[3] Let the equation of Π be given by

$$\sum_{i=1}^{m} \xi_i g_i = \sum_{i=1}^{m} \xi_i c_i .$$

Since N lies entirely on one side of Π, either $\sum_{i=1}^{m} \xi_i g_i \geqq \sum_{i=1}^{m} \xi_i c_i$ for all points $(g_1, \cdots, g_m)$ in N, or $\sum_{i=1}^{m} \xi_i g_i \leqq \sum_{i=1}^{m} \xi_i c_i$ for all $(g_1, \cdots, g_m)$ in N. We put $\xi^1 = -\xi$ if $\Sigma \xi_i g_i \geqq \Sigma \xi_i c_i$ for all points $(g_1, \cdots, g_m)$ in N. Otherwise, we put $\xi^1 = \xi$. Clearly, ξ^1 is maximal relative to c. If ξ^1 is not a complete maximal set relative to c, there exists a vector ξ^2 such that ξ^2 is linearly independent of

ξ^1 and (ξ^1, ξ^2) is maximal relative to c. If (ξ^1, ξ^2) is not a complete maximal set, we can find a vector ξ^3 such that ξ^3 is linearly independent of (ξ^1, ξ^2) and (ξ^1, ξ^2, ξ^3) is a maximal set relative to c, and so on. Continuing this procedure, we shall arrive at a set $(\xi^1, \cdots, \xi^r)(r \leqq m)$ that is a complete maximal set relative to c. This completes the proof of Lemma 3.1.

LEMMA 3.2. *If* $(\xi^1, \cdots, \xi^r)$ *is a maximal set of vectors relative to* $c = (c_1, \cdots, c_m)$ *and if* $\nu(S) = c$, *then the following two conditions are fulfilled for all* x *(except perhaps on a set of measure zero):*

a) *If* x *is a point in* R *for which* $\sum_{j=1}^{m} \xi_j^i f_j(x) = 0$ *for* $i = 1, 2, \cdots, u - 1$ *and*

$\sum_{j=1}^{m} \xi_j^u f_j(x) > 0$ $(u = 1, 2, \cdots, r)$, *then* $x \in S$.

b) *If* x *is a point of* R *for which* $\sum_{j=1}^{m} \xi_j^i f_j(x) = 0$ *for* $i = 1, 2, \cdots, u - 1$ *and*

$\sum_{j=1}^{m} \xi_j^u f_j(x) < 0$, *then* $x \notin S$.

PROOF. Assume that $(\xi^1, \cdots, \xi^r)$ is maximal relative to c. Then, ξ^1 is maximal relative to c. This implies that for all x (except perhaps on a set of measure zero) the following condition holds: $x \in S$ when $\sum_{j=1}^{m} \xi_j^1 f_j(x) > 0$ and $x \notin S$ when

$\sum_{j=1}^{m} \xi_j^1 f_j(x) < 0$. Thus, conditions (a) and (b) of our lemma must be fulfilled for $u = 1$. We shall now show that if (a) and (b) hold for $u = 1, \cdots, v$ then (a) and (b) must hold also for $u = v + 1$. For this purpose, consider the set R' of all points x for which $\sum_{j=1}^{m} \xi_j^i f_j(x) = 0$ for $i = 1, \cdots, v$. If R is replaced by

R', then ξ^{v+1} is maximal relative to $c' = (c_1', \cdots, c_m')$ where $c_i' = \int_{R'} f_i(x)\, dx$ and $S' = S \cap R'$. Hence, for any x in R' (except perhaps on a set of measure zero) the following condition holds: $x \in S$ when $\sum_{j=1}^{m} \xi_j^{v+1} f_j(x) > 0$ and $x \notin S$ when

$\sum_{j=1}^{m} \xi_j^{v+1} f_j(x) < 0$. But this implies that (a) and (b) hold for $u = v + 1$. This completes the proof of our lemma.

LEMMA 3.3. *Let* $(\xi^1, \cdots, \xi^r)$ *be a complete maximal set of vectors relative to* $c = (c_1, \cdots, c_m)$, *and let* T *be the set of all points* $g = (g_1, \cdots, g_m)$ *of* N *for which* $\sum_{j=1}^{m} \xi_j^i g_j = \sum_{j=1}^{m} \xi_j^i c_j$ *for* $i = 1, 2, \cdots, r$. *Then* T *is a bounded, closed and convex set and* c *is an interior point of* T.

PROOF. Clearly, T is a bounded, closed and convex set. Suppose that c is a boundary point of T. Then there exists a hyperplane Π of dimension $m - 1$ such that Π goes through c, Π contains only boundary points of T and T lies entirely on one side of Π^2. Let the equation of Π be given by

$$\sum_{j=1}^{m} \xi_j g_j = \sum_{j=1}^{m} \xi_j c_j,$$

where ξ is independent of $\xi^1, \cdots, \xi^r$. Since T lies on one side of Π, we have either $\sum_{j=1}^{m} \xi_j g_j \geqq \sum_{j=1}^{m} \xi_j c_j$ for all $g = (g_1, \cdots, g_m)$ in T, or $\sum_{j=1}^{m} \xi_j g_j \leqq \sum_{j=1}^{m} \xi_j c_j$ for all g in T. Let $\xi_j^{r+1} = \xi_j (j = 1, \cdots, m)$ in the latter case, and $\xi_j^{r+1} = -\xi_j$ in the former case. Then $\sum_{j=1}^{m} \xi_j^{r+1} g_j \leqq \sum_{j=1}^{m} \xi_j^{r+1} c_j$ for all g in T. But then $(\xi^1, \cdots, \xi^r, \xi^{r+1})$ is a maximal set relative to c, contrary to our assumption that $(\xi^1, \cdots, \xi^r)$ is a complete maximal set. Thus, c must be an interior point of T and our lemma is proved.

THEOREM 3.2. *If $c = (c_1, \cdots, c_m)$ is a boundary point of N and if $(\xi^1, \cdots, \xi^r)$ is a complete maximal set of vectors relative to c, then a necessary and sufficient condition for a member S of S to be a member of S_0 is that there exist m constants $k_1, \cdots, k_m$ such that for all x in R' (except perhaps on a set of measure zero) the inequalities (1.3) and (1.4) hold, where R' is the set of all points x for which*

$$\sum_{j=1}^{m} \xi_j^i f_j(x) = 0 \quad for \quad i = 1, 2, \cdots, r.$$

PROOF. Suppose that $c = (c_1, \cdots, c_m)$ is a boundary point of N and that $(\xi^1, \cdots, \xi^r)$ is a complete maximal set of vectors relative to c. Let R^* be the set of all points x for which the following two conditions hold: (1) $\sum_{j=1}^{m} \xi_j^i f_j(x) \neq 0$ for at least one value i; (2) $\sum_{j=1}^{m} \xi_j^i f_j(x) > 0$ where i is the smallest integer for which $\sum_{j=1}^{m} \xi_j^i f_j(x) \neq 0$. For any member S of S let S^* denote the intersection of S with $R - R'$. It follows from Lemma 3.2 that $R^* - R^* \cap S^*$ and $S^* - R^* \cap S^*$ are sets of measure zero. Thus

$$(3.7) \qquad \int_{S^*} f_i(x) \, dx = \int_{R^*} f_i(x) \, dx \qquad (i = 1, \cdots, m + 1)$$

for all $S \in \mathsf{S}$. Let

$$(3.8) \qquad f_i^*(x) = f_i(x) \quad for \quad x \in R' \qquad (i = 1, \cdots, m + 1)$$

and

$$(3.9) \qquad f_i^*(x) = 0 \quad for \quad x \in R - R' \quad (i = 1, 2, \cdots, m + 1).$$

Let, furthermore,

$$(3.10) \qquad c_i^* = c_i - \int_{R^*} f_i(x) \, dx \qquad (i = 1, \cdots, m).$$

Let $\mu^*, \nu^*, M^*, N^*, \mathsf{S}^*$ and S_0^* have the same meaning with reference to the functions $f_1^*(x), \cdots, f_{m+1}^*(x)$ and the point $c^* = (c_1^*, \cdots, c_m^*)$ as $\mu, \nu, M, N, \mathsf{S}$ and S_0 have with reference to the functions $f_1(x), \cdots, f_{m+1}(x)$ and the point $c = (c_1, \cdots, c_m)$.

It follows from Lemma 3.2 that for any subset S of R for which $\nu(S)$ is a point of the set T defined in Lemma 3.3 we have

$$\int_S f_i(x)\ dx = \int_S f_i^*(x)\ dx + \int_{R'} f_i(x)\ dx \qquad (i = 1, \cdots, m + 1).$$

Since the range of $\nu^*(S)$ is equal to N^* even when S is restricted to subsets S for which $\nu(S) \in T$, the set N^* is obtained from the set T by a translation. The same translation brings the point $c = (c_1, \cdots, c_m)$ into $c^* = (c_1^*, \cdots, c_m^*)$. It then follows from Lemma 3.3 that c^* is an interior point of N^*. Application of Theorem 3.1 gives the following necessary and sufficient condition for a member S of $\mathcal{S}^*$ to be a member of $\mathcal{S}_0^*$: There exist m constants $k_1, \cdots, k_m$ such that for all x (except perhaps on a set of measure zero)

$$(3.11) \qquad f_{m+1}^*(x) \geqq k_1 f_1^*(x) + \cdots + k_m f_m^*(x) \quad \text{when} \quad x \in S$$

and

$$(3.12) \qquad f_{m+1}^*(x) \leqq k_1 f_1^*(x) + \cdots + k_m f_m^*(x) \quad \text{when} \quad x \notin S.$$

It follows from (3.8) and (3.9) that (3.11) and (3.12) are equivalent to

$$(3.13) \qquad f_{m+1}(x) \geqq k_1 f_1(x) + \cdots + k_m f_m(x) \quad \text{when} \quad x \in S \cap R'$$

and

$$(3.14)\ f_{m+1}(x) \leqq k_1 f_1(x) + \cdots + k_m f_m(x) \quad \text{when} \quad x \in (R - S) \cap R'.$$

Theorem 3.2 follows from this and the fact that every member S of $\mathcal{S}$ is a member of $\mathcal{S}^*$ and that a member S of $\mathcal{S}$ is a member of $\mathcal{S}_0^*$ if and only if S is a member of $\mathcal{S}_0$.

It may be of interest to note that if the set R' is of measure zero, the members of $\mathcal{S}$ can differ from each other only by sets of measure zero; i.e., $\mathcal{S}$ consists essentially of one element. This is an immediate consequence of Lemma 3.2.

REFERENCES

[1] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1–37.
[2] A. LYAPUNOV, "Sur les fonctions-vecteurs complètement additives," *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, Vol. 4 (1940), pp. 465–78.
[3] T. BONNESEN AND W. FENCHEL, *Theorie der Konvexen Körper*, Chelsea Publishing Company, New York, 1948.
[4] P. R. HALMOS, "The range of a vector measure," *Bull. Am. Math. Soc.*, Vol. 54 (1948), pp. 416–421.

ESTIMATORS OF THE PROBABILITY OF THE ZERO CLASS IN POISSON AND CERTAIN RELATED POPULATIONS

By N. L. Johnson

University College, London

1. Summary and conclusions. Two estimators of the probability of falling into the zero class are compared, for a family of populations related to Poisson populations. The first estimator, ϵ_1, is based on the observed proportion in the zero class; the second, ϵ_2, would be the maximum likelihood estimator if the underlying distribution were Poisson.

From a practical point of view each estimator possesses its own peculiar advantages. ϵ_1 has the advantage that the detailed distribution among the non-zero classes need not be examined. ϵ_2 has the advantage that only the mean of the observations is needed, the distribution among the various classes not being required. The relative importance of these advantages will naturally vary according to the situations in which the estimators are to be used.

An arbitrary measure of relative accuracy, the mean square error ratio, is used. On this basis ϵ_2 is superior to ϵ_1 for all sample sizes (greater than one) if the population distribution is Poisson. Provided the sample size is not too large ϵ_2 may still be superior to ϵ_1 when the population distribution deviates to a moderate extent from Poisson form.

A third estimator ϵ_3, which is a modification of ϵ_2 and is unbiased, provided the population is Poisson, may be preferred to ϵ_2 unless p exceeds about 0.45. Its properties *vis-á-vis* ϵ_1 probably differ little from those of ϵ_2.

2. The problem. The following investigation was suggested by a problem which arose frequently in connection with the study of weapon lethality in the course of wartime operational and development research. When a fragmenting shell or bomb bursts at a given distance from a target, the density of strikes will vary according to the angular direction with regard to the equatorial plane of the shell. Within the main fragment belt, however, the density may be regarded as varying locally in a random way about an average value. The practical requirement is to determine the chance, say q, that at least one potentially lethal or effective fragment will strike an area of given size which we may call the 'unit area'. Alternatively we can estimate $p = 1 - q$, the chance that no such fragment will strike the unit area.

If it is assumed that the distribution of effective hits follows the Poisson law, and in certain cases evidence indicated that this was justifiable, then $q = 1 - e^{-m}$ and $p = e^{-m}$, where m is the expected value of the number of strikes on the unit area. It was therefore customary to estimate m from the observed average number of effective hits, $\bar{v}$ say, per unit area, derived from a series of experimental firings. Then q was estimated by the formula $1 - e^{-\bar{v}}$. If the distribution

departs from the Poisson form, the procedure is clearly incorrect in theory, but in practice the data were often inadequate to establish any alternative form of the distribution law and the estimator $1 - e^{-\bar{v}}$ was still used. In the discussion below we shall be concerned with the relative accuracy of two alternative estimators of $p(= 1 - q)$ (one of the estimators being $e^{-\bar{v}}$),

(a) when the distribution follows the Poisson law;

(b) when it departs from this law, but can be represented by a positive or negative binomial.

3. Properties of the two estimators. The problem may be stated formally as follows: v_1, v_2, $\cdots$, v_n are independent discrete random variables. If n_0 be the number of zero values out of the n values then

$$(1) \qquad \epsilon_1 = n_0/n$$

may be used as an estimator of p, the probability of the zero class. ϵ_1 is, in fact, the usual form of estimator for the proportion of individuals falling into a given class, and is of general application.

The estimator of p described in section 2 is

$$(2) \qquad \epsilon_2 = e^{-\bar{v}},$$

where $\bar{v} = n^{-1} \sum_{i=1}^{n} v_i$. This estimator is based on the assumption of a common Poisson distribution for the v's.

It will be noted that, while the evaluation of the estimator ϵ_2 does not require a knowledge of the values of the separate v's (provided their total or average is known), ϵ_1 requires only a knowledge of the number of v's which are zero. In the case described in section 2, ϵ_2 is often appropriate as the separate values of the v's are not known though their total is known. On the other hand, if, for example, v_1, v_2, $\cdots$, v_n represent the number of cells developing in a given time in a number of cultures, it may be possible to observe only n_0, the number of cases where no development has occurred. In such cases Fisher [1] has considered the inverse problem of estimating m from n_0 by the formula $-\log \epsilon_1$. This problem will not be considered in the present paper.

We shall now compare the estimators ϵ_1 and ϵ_2 in the case when the v's do, in fact, each follow a Poisson distribution with expected value m, so that

$$(3) \qquad Pr.\{v = r\} = \frac{m^r}{r!} e^{-m} \qquad (r = 0, 1, 2, \cdots).$$

The probability of the zero class is

$$(4) \qquad p = Pr.\{v = 0\} = e^{-m}.$$

Since n_0 is a binomial variable with probability p and index n, the moments and moment-ratios of ϵ_1 are easily determined. In regard to ϵ_2, it can be shown that

$$(5) \qquad \mu_s'(\epsilon_2) = p^{n f(s, n)},$$

where

$$(6) \qquad f(s, n) = 1 - e^{-s/n}.$$

ϵ_1 is an unbiased estimator of p while ϵ_2 is biased. Numerical calculation shows that this bias is negligible for most practical purposes (the maximum absolute bias is in the range $p = 0.3$–0.4 and is approximately $+0.18/n$). For all values of p the relation .

$$(7) \qquad \lim_{n \to \infty} \mathcal{E}(\epsilon_2) = p$$

holds.

4. Comparison of the estimators. Since ϵ_2 is a biased estimator of p, the comparison of ϵ_1 and ϵ_2 certainly cannot be based simply on their variances. One method of comparison, which does make some allowance for biases, is to use the

TABLE I

Ratio of mean square error of ϵ_2 to mean square error of ϵ_1 (Poisson population)

p \ n	10	20	30	60	∞
0.1	0.337	0.296	0.282	0.269	0.256
0.2	0.475	0.439	0.427	0.416	0.402
0.3	0.570	0.544	0.535	0.527	0.516
0.4	0.644	0.628	0.623	0.619	0.611
0.5	0.704	0.700	0.698	0.696	0.693
0.6	0.756	0.762	0.763	0.767	0.766
0.7	0.800	0.816	0.822	0.829	0.832
0.8	0.839	0.866	0.875	0.886	0.893
0.9	0.874	0.911	0.923	0.938	0.948

mean square errors of the estimators [2]. The mean square error of ϵ_2 is $\mathcal{E}[(\epsilon_2 - p)^2]$ $= \sigma^2(\epsilon_2) + [\mathcal{E}(\epsilon_2) - p]^2$, while the mean square error of ϵ_1 is $\mathcal{E}[(\epsilon_1 - p)^2] = \sigma^2(\epsilon_1)$ since ϵ_1 is an unbiased estimator of p. The ratio of mean square errors will be used as an index of comparison of estimators in the present paper, although it is clearly arbitrary, and other criteria could be preferable in certain circumstances.

Table I gives values of the mean square error ratio for various values of n and p. According to this criterion the second estimator (ϵ_2) is more accurate than the first (ϵ_1) for all cases shown in this table.

It can be shown that this ratio of mean squares must always be less than one, except in the trivial case $n = 1$. The relative advantage of ϵ_2 increases as p diminishes and does not vary greatly with n.

The correlation between the two estimators is

$$(8) \qquad \rho(\epsilon_1, \epsilon_2) = (np)^{\frac{1}{2}}(1 - p)^{-\frac{1}{2}}\{p^{-f(1,n)} - 1\} \{p^{-n[f(1,n)]^2} - 1\}^{-\frac{1}{2}},$$

whence

(9)
$$\lim_{n \to \infty} \rho(\epsilon_1, \epsilon_2) = \{-p(1-p)^{-1} \log p\}^{\frac{1}{2}}.$$

$\rho(\epsilon_1, \epsilon_2)$ approaches this limit rapidly as n increases. We note that

(10)
$$\lim_{n \to \infty} \rho(\epsilon_1, \epsilon_2) = \lim_{n \to \infty} (\sigma(\epsilon_2)/\sigma(\epsilon_1)),$$

as is to be expected since ϵ_2 is the maximum likelihood estimator of p [3].

5. A third estimator of p. The superiority of ϵ_2 as an estimator of p is to be expected, since $\bar{v}$ is a sufficient statistic for p. Using the method described in [4], we obtain the minimum variance unbiased estimator[1]

(11)
$$\epsilon_3 = (1 - n^{-1})^{n\bar{v}},$$

which may be regarded as a modified, and perhaps improved, form of ϵ_2.

The variance of ϵ_3 is $p^2(p^{-1/n} - 1)$. This differs but little from the mean square error of ϵ_2, as is to be expected since $(1 - n^{-1})^n \doteq e^{-1}$. It appears that for sufficiently large values of n the mean square error of ϵ_3 will be slightly less than that of ϵ_2 for $p < 0.45$, while for $p > 0.45$ the mean square error of ϵ_2 will be slightly the smaller. The performance of ϵ_3 compared with ϵ_1 will be practically identical with that of ϵ_2.

6. Non-Poisson populations. It is quite possible that ϵ_2 (or ϵ_3) may be used as an estimator of p even when v is not in fact a Poisson variable. It may be that it has been incorrectly assumed that the distribution is Poisson in form or, perhaps, departure from Poisson, though admitted, has been considered of insufficient magnitude to affect the usefulness of ϵ_2.

It is of interest to investigate the effect of deviations from the Poisson distribution on the properties of ϵ_1 and ϵ_2. In order to do this it is first necessary to specify the nature of these deviations. Many forms of modification of the Poisson distribution have been suggested ([5]-[9]). We shall deal only with the simple form of deviation from Poisson wherein the distribution is defined by successive terms in the expansion of

(12)
$$[(1 + \omega) - \omega]^{-m/\omega}, \qquad -1 < \omega < 0 \text{ or } 0 < \omega.$$

The expected value of this distribution is m, whatever be the value of ω. If $-1 < \omega < 0$, then putting $\omega = -P, 1 + \omega = Q, NP = m$ we have the binomial distribution

(13)
$$Pr\{v = r\} = \binom{N}{r} P^r Q^{N-r}.$$

---

[1] I am indebted to the referee for suggesting the use of this estimator.

If $0 < \omega$ we have the negative binomial distribution. Putting $\omega = 2\sigma^2$, $m = f\sigma^2$ we have

$$(14) \qquad Pr\{v = r\} = \frac{\Gamma(r + \frac{1}{2}f)}{r!\Gamma(\frac{1}{2}f)} - \frac{(2\sigma^2)^r}{(2\sigma^2 + 1)^{r+\frac{1}{2}f}},$$

a form of the Pólya-Eggenberger [10] distribution previously obtained by Greenwood & Yule [11], which can be considered to arise from a mixture of Poisson distributions with expected values distributed proportionately to $\chi^2\sigma^2$ with f degrees of freedom. As $\omega \to 0$, the distribution tends to the Poisson form whether ω is moving through positive or negative values.

Whether ω is positive or negative, the probability of the zero class is

$$(15) \qquad p = (1 + \omega)^{-m/\omega}.$$

The moments and moment-ratios of ϵ_1 are the same functions of p as in the Poisson case. It can be shown that

$$(16) \qquad \mu_s'(\epsilon_2) = [1 + \omega f(s, n)]^{-mn/\omega},$$

where $f(s, n) = 1 - e^{-s/n}$ as in (6), and that the correlation between the two estimators is

$$(17) \qquad \rho(\epsilon_1, \epsilon_2) = (np)^{\frac{1}{2}}(1 - p)^{-1}\{[1 + \omega f(1, n)]^{m/\omega} - 1\} \\ \cdot \{[1 + \omega f(2, n)]^{-mn/\omega}[1 + \omega f(1, n)]^{2mn/\omega} - 1\}^{-\frac{1}{2}}.$$

For any value of p, ϵ_1 is still an unbiased estimator of p, and has the same variance as when the distribution of v is Poisson. ϵ_2 is still a biased estimator of p, but the amount of bias and the variance of ϵ_2 are not the same as when the distribution of v is Poisson. Furthermore (7) no longer holds. In fact, putting $s = 1$ in (16)

$$(18) \qquad \begin{aligned} \mathcal{B}(\epsilon_2) &= [1 + \omega(1 - e^{-1/n})]^{-mn/\omega}, \\ \lim_{n \to \infty} \mathcal{B}(\epsilon_2) &= p^{\omega/\log(1+\omega)} \neq p. \end{aligned}$$

7. Approximations. Since the formulae in (16) and (17) are tedious to compute, it seemed worth while investigating whether any simple approximations were possible. The following expansions in powers of n^{-1} up to the term in n^{-1} were found to give generally good results for $n \geq 30$.

$$(19.1) \qquad \mathcal{B}(\epsilon_2) \fallingdotseq e^{-m}[1 + \frac{1}{2}m(1 + \omega)n^{-1}],$$

$$(19.2) \qquad \sigma^2(\epsilon_2) \fallingdotseq e^{-2m}m(1 + \omega)n^{-1},$$

$$(19.3) \qquad \sqrt{\beta_1}(\epsilon_2) \fallingdotseq [nm(1 + \omega)]^{-\frac{1}{2}}[3m(1 + \omega) - (1 + 2\omega)],$$

$$(19.4) \qquad \beta_2(\epsilon_2) \fallingdotseq 3 + 16[nm(1 + \omega)]^{-1}[m^2(1 + \omega)^2 - 12m(1 + \omega) \\ \cdot (1 + 2\omega) + 1 + 6\omega + 6\omega^2],$$

$$(19.5) \qquad \rho(\epsilon_1, \epsilon_2) \fallingdotseq (-\omega p \log p)^{\frac{1}{2}}[(1 + \omega)(1 - p) \log (1 + \omega)]^{-\frac{1}{2}} \\ \cdot [1 + (\frac{1}{4}m + \frac{1}{2}\omega - \frac{1}{4}m\omega)n^{-1}].$$

The values of $\mathcal{E}(\epsilon_2)$ and $\sigma^2(\epsilon_2)$ obtained from the exact formula (16) and from (19) are compared in Tables II and III respectively.

It should be noted that some of the values of ω shown do not correspond to real distributions. These cases are indicated by parentheses enclosing the corresponding figures. The values of ω chosen exhibit the trend of mathematical

TABLE II

Expected value of ϵ_2

(*Note*: The exact values and (19.1) agree to three decimal places for all cases included in this table.)

p	ω	$n = 30$	$n = 60$	$n = \infty$
0.1	−0.50	(0.193)	(0.191)	(0.190)
	−0.25	(0.139)	(0.137)	(0.135)
	0.00	0.104	0.102	0.100
	1.00	0.040	0.038	0.036
0.5	−0.25	(0.552)	(0.550)	(0.548)
	0.00	0.506	0.503	0.500
	1.00	0.380	0.374	0.368
0.9	0.00	0.902	0.901	0.900
	1.00	0.863	0.861	0.859

TABLE III

Approximate and exact values of $100 \, \sigma^2(\epsilon_2)$

p	ω	$n = 30$		$n = 60$	
		Approx.	Exact	Approx.	Exact
0.1	−0.50	(0.100)	(0.104)	(0.050)	(0.051)
	−0.25	(0.091)	(0.097)	(0.046)	(0.047)
	0.00	0.077	0.083	0.038	0.040
	1.00	0.029	0.036	0.014	0.016
0.5	−0.25	(0.451)	(0.454)	(0.226)	(0.226)
	0.00	0.578	0.578	0.289	0.289
	1.00	0.901	0.910	0.451	0.451
0.9	0.00	0.284	0.277	0.142	0.140
	1.00	0.748	0.689	0.374	0.359

functions of ω which do give the moments of ϵ_2 for real distributions when ω takes certain special values, different for different p. The functions are simple continuous functions of ω and the method of presentation should not prove misleading.

Close agreement was also obtained between values given by (19.3)–(19.5) and the corresponding exact values. The approximation to $\sqrt{\beta_1}(\epsilon_2)$ was generally

correct to two decimal places and that to $\rho(\epsilon_1, \epsilon_2)$ was generally correct to three places for the values of n, ω and p in Tables II and III. $\beta_2(\epsilon_2)$ was correct to two decimal places for ω negative, while for positive ω the error did not exceed 0.04 except for $p = 0.1$ and $\omega = 1.0$ (5.09 (approx.) against 5.46).

TABLE IV
Values of $n(\omega,p)$

p \\ ω	-0.2	-0.1	$+0.1$	$+0.2$
0.1	80	400	620	190
0.5	70	270	250	60
0.9	270	680	*	*

* Formula (21) gives negative values in these cases.

TABLE V

p	$f_{-2}(p)$	$f_{-1}(p)$	$f_0(p)$
0.1	5.0528	10.8992	7.5954
0.2	3.6916	6.4732	5.1006
0.3	3.1164	3.9314	3.8761
0.4	2.7809	1.6529	2.7640
0.5	2.5547	-0.9192	1.4261
0.6	2.3889	-4.3392	-0.4658
0.7	2.2606	-9.6654	-3.5511
0.8	2.1574	-19.9286	-9.6719
0.9	2.0722	-50.1476	-28.0060

8. A critical sample size. Using the approximate formulae (19) we see that the mean square error of ϵ_1 will be less than that of ϵ_2 provided

$$(20) \quad p(1-p)n^{-1} < (e^{-m} - p)^2 + m(1+\omega)\{e^{-m}(e^{-m} - p) + e^{-2m}\}n^{-1}.$$

This can be rewritten $n > n(\omega, p)$, where

$$(21) \quad n(\omega, p) = [p(1-p) - m(1+\omega)e^{-m}(2e^{-m} - p)](e^{-m} - p)^{-2}.$$

Provided the value of $n(\omega, p)$ given by (21) is sufficiently large for the approximation in (19) to be good, it can be said that ϵ_1 will be a better estimator of p than ϵ_2 (according to the mean square error criterion) if the sample size is bigger than $n(\omega, p)$. For smaller sample sizes it is likely that ϵ_2 will still be the superior estimator as in the Poisson case.

When $|\omega|$ is small the expansion

$$(22) \quad n(\omega, p) = \omega^{-2}f_{-2}(p) + \omega^{-1}f_{-1}(p) + f_0(p) + \cdots$$

where

$$(23.1) \quad f_{-2}(p) = 4(p \log p)^{-2}[p(1-p) + p^2 \log p],$$

$$(23.2) \quad f_{-1}(p) = 4(p \log p)^{-2}[(\tfrac{1}{3} - \tfrac{1}{2} \log p)p(1 - p) + \left(\frac{11}{6} + \log p\right)p^2 \log p],$$

$$f_0(p) = 4(p \log p)^{-2}\left[\left\{\frac{5}{48} (\log p)^2 - \frac{1}{12} \log p - \frac{1}{12}\right\}\right.$$

$$(23.3)$$

$$\left. \cdot p(1 - p) + \left\{\frac{11}{48} (\log p)^2 + \frac{5}{3} \log p + \frac{5}{6}\right\} p^2 \log p \right],$$

is useful. The values of $n(\omega, p)$ given by the series (22) taken as far as $f_0(p)$ agree (to the nearest ten) with those in Table IV, which were calculated from (21). Values of $f_{-2}(p), f_{-1}(p)$ and $f_0(p)$ for $p = 0.1 - (0.1) - 0.9$ are given in Table V.

REFERENCES

[1] E. Johnson, "Estimates of parameters by means of least squares", *Annals of Math. Stat.*, Vol. 11 (1940), p. 453.

[2] R. A. Fisher, *The Design of Experiments*, 1st ed., Oliver & Boyd, 1935, p. 221.

[3] D. Blackwell, "Conditional expectation and unbiased sequential estimation", *Annals of Math. Stat.*, Vol. 18 (1947), p. 105.

[4] R. A. Fisher, "Theory of statistical estimation", *Proc. Camb. Phil. Soc.*, Vol. 22 (1925), p. 700.

[5] C. V. L. Charlier, *Die Grundzüge der Mathematische Statistik*, Lund. Verlag Scientia, 1920, p. 80.

[6] G. Pólya, "Sur quelques points de la théorie des probabilités", *Ann. de l'Inst. H. Poincaré*, Vol. 1 (1931), p. 117.

[7] J. Neyman, "On a new class of 'contagious' distributions", *Annals of Math. Stat.*, Vol. 10 (1939), p. 35.

[8] W. Feller, "On a general class of 'contagious' distributions", *Annals of Math. Stat.*, Vol. 14 (1943), p. 389.

[9] M. Thomas, "A generalization of Poisson's binomial limit for use in ecology", *Biometrika*, Vol. 36 (1949), p. 18.

[10] F. Eggenberger and G. Pólya, "Über die Statistik verketteter Vorgänge", *Zeit. für Ang. Math. und Mech.*, Vol. 1 (1923), p. 279.

[11] M. Greenwood and G. Yule, "An inquiry into the nature of frequency distributions of multiple happenings", *Jour. Roy. Stat. Soc.*, Vol. 83 (1920), p. 255.

TESTING PROPORTIONALITY OF COVARIANCE MATRICES[1]

By Walter T. Federer

Cornell University

1. Summary. The problem of comparing the proportionality of covariance matrices often arises in genetic experiments. Knowledge of nonproportionality of covariance matrices is useful in selection work and in genetic interpretations. In developing a test of significance for this contrast, the likelihood ratio criterion was used. Likelihood ratio tests were obtained for two sets and for three sets of independent variance-covariance matrices. The test for r independent covariances was indicated and some unsolved problems were cited.

2. Introduction. Tests of significance of variances from normally distributed variates are available for testing the equality of:

(i) Two independent variances (Snedecor's F, Fisher's z, Mahalanobis' x [3], and Fisher and Yates' variance ratio),

(ii) k independent variances (Chi-square tests by Stevens [6], Bartlett [1], and Cochran [2]),

(iii) Two variances with unknown correlation (Pitman [5] and Morgan's test [4] and Wilks' likelihood-ratio test [8]),

(iv) k variances and of the associated covariances (Wilks' likelihood-ratio test [8]),

(v) The variances and covariances within each of several sets and the covariances between sets (Likelihood-ratio tests by Votaw [7]),

but no tests of significance are available for comparing the proportionality of two or more variance-covariance matrices.

The hypothesis of proportionality of variance-covariance matrices is more tenable than equality in many genetic experiments, since it is known that the variances are unequal but it is not known if the variance or covariance for one strain is merely a multiple of that for the other strain. Knowledge of this is of importance in any genetic study on the inheritance of characters and in selection work. In addition, the means and variances are often related in some manner and a transformation of the data may not be advisable since this may lead to incorrect genetic interpretations.

3. Likelihood ratio for comparing two covariance matrices. The problem of testing the hypothesis that the variances and covariances from strain A are proportional to the variances and covariances of strain B was solved by an

---

application of the likelihood-ratio test. Let the characters be represented by $X_1, X_2, \cdots, X_p$ for strain A and by $Y_1, Y_2, \cdots, Y_p$ for strain B, respectively. The hypothesis, then, is that the variance or covariance of A equals K times the corresponding variance or covariance for B, that is, $(\sigma_{ij})_x = K(\sigma_{ij})_y$, where K is a proportionality factor and the sample variance-covariance matrices for A and B are independently estimated.

The likelihood ratio for the above in general terminology is

$$\lambda = f(a_{ij}, b_{ij}, \bar{K}, \hat{\sigma}_y^{ij})/f(a_{ij}, \hat{\sigma}_x^{ij}) f(b_{ij}, \hat{\sigma}_y^{ij}),$$

where

$$a_{ij} = \sum_{u=1}^{n} (X_{iu} - \bar{x}_i)(X_{ju} - \bar{x}_j), \quad b_{ij} = \sum_{u=1}^{m} (Y_{iu} - \bar{y}_i)(Y_{ju} - \bar{y}_j),$$

X_{iu} and Y_{iu} are the sample elements and $\bar{x}_i$ and $\bar{y}_i$ are the sample means, $i, j = 1, 2, \cdots, p$, $\bar{K}$ and $\hat{\sigma}_y^{ij}$ are the maximum likelihood estimates of K and σ_y^{ij} computed under the hypothesis that $(\sigma_{ij})_x = K(\sigma_{ij})_y$, $\hat{\sigma}_x^{ij}$ and $\hat{\sigma}_y^{ij}$ are the maximum likelihood estimates of σ_x^{ij} and σ_y^{ij} computed under the hypothesis of independence, and where there are $n - 1$ degrees of freedom associated with the a_{ij} and $m - 1$ with the b_{ij}.

It is known that the sums of squares and cross products of p normally distributed variates follow the Wishart distribution with $n - 1$ degrees of freedom. Furthermore, the joint distribution of two independent sums of squares and cross products may be written as

$$f(a_{ij}, b_{ij}, \sigma_x^{ij}, \sigma_y^{ij}) = f(a_{ij}, \sigma_x^{ij}) f(b_{ij}, \sigma_y^{ij}),$$

which is proportional to

$$|\sigma_x^{ij}|^{\frac{1}{2}(n-1)} |\sigma_y^{ij}|^{\frac{1}{2}(m-1)} |a_{ij}|^{\frac{1}{2}(n-p-2)} |b_{ij}|^{\frac{1}{2}(m-p-2)} \exp\left[-\tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} (\sigma_x^{ij} a_{ij} + \sigma_y^{ij} b_{ij})\right].$$

The maximum likelihood estimates for σ_x^{ij} and σ_y^{ij} are:

$$\| \hat{\sigma}_x^{ij} \| = \| a_{ij}/(n - 1) \|^{-1} \quad \text{and} \quad \| \hat{\sigma}_y^{ij} \| = \| b_{ij}/(m - 1) \|^{-1};$$

also $(\hat{\sigma}_{ij})_x = a_{ij}/(n - 1)$ and $(\hat{\sigma}_{ij})_y = b_{ij}/(m - 1)$.

Now under the hypothesis that the variances and covariances are proportional, i.e., $(\sigma_{ij})_x = K(\sigma_{ij})_y$, the joint distribution is proportional to

$$|\sigma_y^{ij}/K|^{\frac{1}{2}(n-1)} |\sigma_y^{ij}|^{\frac{1}{2}(m-1)} |a_{ij}|^{\frac{1}{2}(n-p-2)} |b_{ij}|^{\frac{1}{2}(m-p-2)} \exp\left[-\tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma_y^{ij}(a_{ij}/K + b_{ij})\right].$$

The maximum likelihood estimates of K and σ_y^{ij} are obtained from the equations

$$\bar{K} = \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\sigma}_y^{ij} a_{ij}/p(n - 1)$$

and

$$\| \hat\sigma_v^{ij} \| = \| (a_{ij} + \bar{K}b_{ij})/\bar{K}(n + m - 2) \|^{-1}.$$

It is possible to solve for $\bar{K}$ in the above 2 equations. For $p = 2$ the equation in $\bar{K}$ free of $\hat\sigma_v^{ij}$ is

$$\bar{K}^2 \begin{vmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{vmatrix} + \bar{K}\left(1 - \frac{n + m - 2}{2(n - 1)}\right)\left(\begin{vmatrix} a_{11} & a_{12} \\ b_{12} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{12} & a_{22} \end{vmatrix}\right)$$

$$+ \left(1 - \frac{2(n + m - 2)}{2(n - 1)}\right)\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = 0.$$

For $p = 3$, $\bar{K}$ is obtained by solving the following equation:

$$\bar{K}^3 \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{11} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix}$$

$$+ \bar{K}^2\left(1 - \frac{n + m - 2}{3(n - 1)}\right)\left(\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\right)$$

$$+ \bar{K}\left(1 - \frac{2(n + m - 2)}{3(n - 1)}\right)\left(\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\right)$$

$$+ \left(1 - \frac{3(n + m - 2)}{3(n - 1)}\right)\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0.$$

For $p = 4$ and higher, the coefficients are obtained in a like manner. The number of determinants for each power of $\bar{K}$ will be the same as the coefficients in the binomial.

The proof that there is only one positive root in the polynomial in $\bar{K}$ would be obtained by proving that the sum of the determinants, associated with any power of $\bar{K}$, is positive.[2] Since the other coefficients in the polynomial are positive up to a certain point and negative thereafter, there is only one change in sign, and thus only one positive real root. The positive root is the only one of interest here since the variances are inherently positive.

The likelihood ratio for comparing the proportionality of the variances and

---

[2] A proof of this was first called to my attention by Isadore Blumen.

covariances for strains A and B is

$$\lambda = \bar{K}^{\frac{1}{2}p(m-1)} \, |\, a_{ij}/(n - 1)\,|^{\frac{1}{2}(n-1)} \, |\, b_{ij}/(m - 1)\,|^{\frac{1}{2}(m-1)}$$
$$|(\, a_{ij} + \bar{K}b_{ij})/(n + m - 2)\,|^{-\frac{1}{2}(n+m-2)},$$

where $-2 \log \lambda$ is distributed approximately as chi-square with

$$p(p + 1) - \tfrac{1}{2}p(p + 1) - 1 = \tfrac{1}{2}p(p + 1) - 1$$

degrees of freedom when m and n are large.

4. Likelihood ratio for comparing three covariance matrices. In the event that 3 independently estimated sets of variances and covariances are compared for proportionality, the likelihood ratio is

$$\lambda = f(a_{ij}, b_{ij}, c_{ij}, \bar{K}_1, \bar{K}_2, \hat{\sigma}_s^{ij})/f(a_{ij}, \hat{\sigma}_s^{ij})f(b_{ij}, \hat{\sigma}_y^{ij})f(c_{ij}, \hat{\sigma}_s^{ij}),$$

where $c_{ij} = \sum_{w=1}^{q} (Z_{iw} - \bar{z}_i)(Z_{jw} - \bar{z}_j)$, Z_{iw} are the sample elements and $\bar{z}_i$ the sample means, $\bar{K}_1$, $\bar{K}_2$, and $\hat{\sigma}_s^{ij}$ are the maximum likelihood estimates of K_1, K_2, and σ_s^{ij} computed under the hypothesis of proportionality, that is, $(\sigma_{ij})_s = K_1(\sigma_{ij})_y = K_2(\sigma_{ij})_s$, $(\hat{\sigma}_{ij})_s$, $(\hat{\sigma}_{ij})_y$, and $(\hat{\sigma}_{ij})_s$ are the maximum likelihood estimates of $(\sigma_{ij})_s$, $(\sigma_{ij})_y$, and $(\sigma_{ij})_s$ computed under the hypothesis of independence, a_{ij}, b_{ij}, and c_{ij} have $n - 1$, $m - 1$, and $q - 1$ degrees of freedom respectively, and the a_{ij} and b_{ij} are as defined previously.

Under the hypothesis of independence the maximum likelihood estimates of σ_s^{ij}, σ_y^{ij}, and σ_s^{ij} are $\hat{\sigma}_s^{ij}$ and $\hat{\sigma}_y^{ij}$, given in Section 3, and $\hat{\sigma}_s^{ij}$ which can be obtained from the equation $||\, \hat{\sigma}_s^{ij}\, || = ||\, c_{ij}/(q - 1)\, ||^{-1}$.

Under the hypothesis of proportionality the maximum likelihood estimates of K_1, K_2, and σ_s^{ij} are obtained from the equations:

$$\bar{K}_2 = \bar{K}_1 \Sigma\Sigma\hat{\sigma}_s^{ij}b_{ij}/p(m - 1),$$
$$\bar{K}_2 = \Sigma\Sigma\hat{\sigma}_s^{ij}(a_{ij} + \bar{K}_1 b_{ij})/p(n + m - 2)$$

and

$$||\, \hat{\sigma}_s^{ij}\, || = ||\, (a_{ij} + \bar{K}_1 b_{ij} + \bar{K}_2 c_{ij})/\bar{K}_2(m + n + q - 3)\, ||^{-1}.$$

The positive roots (probably only one for each proportionality constant) for $\bar{K}_1$ and $\bar{K}_2$ which maximize the likelihood ratio are the ones used. Substituting these values in the likelihood ratio the following results:

$$\lambda = \bar{K}_2^{-\frac{1}{2}p(m+n-2)} \, \bar{K}_1^{\frac{1}{2}p(m-1)} \, |\, a_{ij}/(n - 1)\,|^{\frac{1}{2}(n-1)} \, |\, b_{ij}/(m - 1)\,|^{\frac{1}{2}(m-1)}$$
$$|\, c_{ij}/(q - 1)\,|^{\frac{1}{2}(q-1)} \, |\, \hat{\sigma}_s^{ij}\,|^{\frac{1}{2}(m+n+q-3)}.$$

When sample sizes are large, $-2 \log \lambda$ is distributed approximately as chi-square with $p(p + 1) - 2$ degrees of freedom.

The method for comparing r independent sets of variances and covariances

follows by a simple extension of the above likelihood ratio and the solution for the $r - 1$ proportionality constants.

5. Unsolved problems. The nature of the roots for the proportionality constants requires further study. Also, likelihood ratios could be developed for comparing the proportionality of r non-independent covariance matrices under various hypotheses. A study of these tests of significance could be made in much the same way as described by Votaw [7]. Such a study is necessary before a complete understanding of this test is obtained.

REFERENCES

[1] M. S. BARTLETT, "Some examples of statistical methods of research in agriculture and applied biology," *Jour. Roy. Stat. Soc., Suppl.*, Vol. 4 (1937), pp. 137–183.

[2] W. G. COCHRAN, "Problems arising in the analysis of a series of similar experiments," *Jour. Roy. Stat. Soc., Suppl.*, Vol. 4 (1937), pp. 102–118.

[3] P. C. MAHALANOBIS, "Statistical notes for agricultural workers. No. 3—Auxiliary tables for Fisher's z-test in analysis of variance," *Ind. Jour. Agri. Sci.*, Vol. 2 (1932), pp. 679–693.

[4] W. A. MORGAN, "A test for the significance of the difference between the two variances in a sample from a normal bivariate population," *Biometrika*, Vol. 31 (1939), pp. 13–19.

[5] E. J. G. PITMAN, "A note on normal correlation," *Biometrika*, Vol. 31 (1939), pp. 9–12.

[6] W. L. STEVENS, "Heterogeneity of a set of variances," *Jour. of Genetics*, Vol. 33 (1936), pp. 398–399.

[7] D. F. VOTAW, JR., "Testing compound symmetry in a normal multivariate distribution," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 447–473.

[8] S. S. WILKS, "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 257–281.

AN INVERSE MATRIX ADJUSTMENT ARISING IN DISCRIMINANT ANALYSIS

By M. S. Bartlett

University of Manchester, England

1. Introduction. The adjustment of an inverse matrix arising from the change of a single element, or of elements in a single row or column, in the original matrix has recently been discussed by Sherman and Morrison [1, 2]. In discriminant function analysis the adjustment due to the addition of a degenerate matrix of rank one to the original matrix has sometimes been required, and the method used by the writer is described in this note. It will be noticed that this case includes the cases considered by Sherman and Morrison.

2. General formula. The new square matrix can always be written in the form

$$(1) \qquad \mathbf{B} = \mathbf{A} + \mathbf{u}\mathbf{v}',$$

where $\mathbf{u}$ is a column vector (single column matrix), and $\mathbf{v}'$ a row vector (dashes denote matrix transposes). We write formally

$$
\begin{aligned}
\mathbf{B}^{-1} &= (\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1}(1 + \mathbf{u}\mathbf{v}'\mathbf{A}^{-1})^{-1} \\
&= \mathbf{A}^{-1}(1 - \mathbf{u}\mathbf{v}'\mathbf{A}^{-1} + \mathbf{u}\mathbf{v}'\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1} - \cdots) \\
(2) \qquad &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}\,\{1 - \mathbf{v}'\mathbf{A}^{-1}\mathbf{u} + (\mathbf{v}'\mathbf{A}^{-1}\mathbf{u})^2 - \cdots\} \\
&= \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\cdot\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}},
\end{aligned}
$$

which has the same simple structure as (1) and can be determined when $\mathbf{A}^{-1}$ is known. To check this formal result, we may easily verify that pre- or post-multiplication of the expression (2) by $\mathbf{B}$ gives the unit matrix.

3. Numerical example in discriminant analysis. The general regression relation between two sample matrices S_2 and S_1 may be written (Bartlett [3])

$$(3) \qquad S_2 = C_{21}C_{11}^{-1}S_1 + S_{2.1}.$$

Here the n observations of any variable (measured if necessary from the general mean) comprise one row in the appropriate matrix, S_2 and S_1 representing respectively the dependent and independent variables. S_2S_1' is written C_{21} for convenience, and similarly for C_{11}, C_{22}; also $C_{22.1} = S_{2.1}S_{2.1}'$. In discriminant analysis in its strict sense S_1 stands for a single dummy variable serving to isolate a group or other contrast between the proper random variables S_2'. In that case the equation

$$(4) \qquad C_{22} = C_{21}C_{11}^{-1}C_{12} + C_{22.1}$$

107

derived from (3) becomes of the form

(5) $$C_{22} = zz' + C_{22.1}.$$

The discriminant function coefficients in Fisher's original discussion [4] of this type of analysis are proportional to the solution a of the equation

(6) $$C_{22.1}a = z$$

(see Bartlett [3], p. 37), and hence are obtained as $C_{22.1}^{-1}z$, where $C_{22.1}$ is the matrix of 'sums of squares and products' within groups. But in tests of significance of a it is convenient (see, for example, Bartlett [5], §5) to make use of the 'inverted regression relation' (first noted by Fisher [4], p. 184)

(7) $$S_1 = C_{12}C_{22}^{-1}S_2 + S_{1.2},$$

giving discriminant function coefficients $b = C_{22}^{-1}C_{21}$.

It is sometimes required to obtain the second (equivalent) form of solution involving C_{22}^{-1} from computations already available based on the first method of analysis involving $C_{22.1}^{-1}$. For example, in Fisher's original comparison of *Iris versicolor* and *Iris setosa* based on 50 observations, on each species, of the variables

$$x_1 = \text{sepal length,}$$
$$x_2 = \text{sepal width,}$$
$$x_3 = \text{petal length,}$$
$$x_4 = \text{petal width,}$$

he gives (p. 181) for $C_{22.1}^{-1}$ the (symmetric positive definite) matrix

		x_1	x_2	x_3	x_4
	x_1	0.1187161			
(8)	x_2	−0.0668666	0.1452736		
	x_3	−0.0816158	+0.0334101	0.2193614	
	x_4	+0.0396350	−0.1107529	−0.2720206	0.8945506.

We take S_1 as a pseudo-variate with value $+\frac{1}{2}$ for one species and $-\frac{1}{2}$ for the other, so that $C_{11} = 25$, and C_{21} is the column vector of differences in means multiplied by 25, and $z' = C_{21}/\sqrt{25}$. From (5) and (2) the inverse of C_{22} is

$$C_{22.1}^{-1} - \frac{C_{22.1}^{-1}z \cdot z'C_{22.1}^{-1}}{1 + z'C_{22.1}^{-1}z},$$

or from (6),

(9) $$C_{22.1}^{-1} - \frac{aa'}{1 + a'z}.$$

Fisher actually gives the solution of (6) with z replaced by the vector of mean differences, so that, in terms of his solution c, where

(10) $$c = a/5 = \begin{pmatrix} -0.0311511 \\ -0.1839075 \\ +0.2221044 \\ +0.3147370 \end{pmatrix},$$

we find that (9) becomes

(11) $$C_{22.1}^{-1} - 0.9146\ cc'.$$

Hence we obtain C_{22}^{-1} (without having to re-work it from C_{22}) as

		x_1	x_2	x_3	x_4
(12)	x_1	0.11783			
	x_2	−0.07211	0.11434		
	x_3	−0.07529	+0.07077	0.17424	
	x_4	+0.04860	−0.05781	−0.33595	0.80395.

With this matrix we can complete the formal regression analysis of S_1, giving for b and its 'standard errors'

(13)
$$
\begin{aligned}
-0.02847 &\pm 0.03368 \\
-0.16808 &\pm 0.03318 \\
+0.20298 &\pm 0.04095 \\
+0.28764 &\pm 0.08798.
\end{aligned}
$$

The solution b we know to be a multiple of the solution c (as may be verified to within 2 in the fourth decimal place), but we also see from (12) that the first variable is not contributing to the discrimination and might be omitted. The corresponding analysis of variance of S_1 (c.f. Fisher's Table VII) gives

		D.F.	S.S.	M.S.
(14)	between { x_2, x_3, x_4...............	3	24.0785	
	species { x_1 (partial).............	1	0.0069	
	within species......................	95	0.9146	0.011088
	Total............................	99	25.0000	

so that the square of the multiple correlation coefficient is only reduced from 0.96342 to 0.96314 by the omission of x_1. It should be noticed that the multiplier 0.9146 in (11) is the 'within species' entry in (14).

4. Theoretical example in discriminant analysis. The formula (2) is also theoretically useful in deriving the discriminant function by 'size and shape' suggested by Penrose [6]. It is known that for multivariate normal variables x with constant variance matrix V the ideal discriminant function for contrasting two groups has coefficients $d'V^{-1}$, where d is the column vector of true differences in means of the two groups. It is now assumed that after standardization of each variable to unit variance we can write

(15)
$$
V = \begin{bmatrix}
1 & \rho & \rho & \cdots \\
\rho & 1 & \rho & \cdots \\
\cdot & & & \\
\cdot & & & \\
\cdot & & &
\end{bmatrix}
= (1 - \rho)I + \rho ww',
$$

where I is the unit matrix and $\dot{\mathbf{w}}$ a column vector with unit components. Applying formula (2), we find the inverse matrix

$$(16) \qquad \mathbf{V}^{-1} = \frac{\mathbf{I}}{1-\rho} - \frac{\rho}{1-\rho}\frac{\mathbf{w}\mathbf{w}'}{1+\rho(p-1)},$$

where p is the number of variables. Hence

$$\mathbf{d}'\mathbf{V}^{-1} = \frac{\mathbf{d}'}{1-\rho} - \frac{\rho}{1-\rho}\frac{(\mathbf{w}'\mathbf{d})\mathbf{w}'}{1+\rho(p-1)}$$

$$(17) \qquad = \frac{\mathbf{w}'\mathbf{d}}{p(1-\rho)}\left\{\left[\frac{p\mathbf{d}'}{\mathbf{w}'\mathbf{d}} - \mathbf{w}'\right] + \mathbf{w}'\left[1 - \frac{p\rho}{1+\rho(p-1)}\right]\right\}$$

$$\propto \left[\frac{p\mathbf{d}'}{\mathbf{w}'\mathbf{d}} - \mathbf{w}'\right] + \mathbf{w}'\left[\frac{1-\rho}{1+\rho(p-1)}\right],$$

where the two sets of coefficients in (17), $\mathbf{h}'$ and $\mathbf{g}'$ ($\propto \mathbf{w}'$), say (respectively), are arranged to give zero correlation between $\mathbf{g}'\mathbf{x}$ and $\mathbf{h}'\mathbf{x}$. This is checked by evaluating the covariance $E\{\mathbf{w}'\mathbf{y}\cdot\mathbf{h}'\mathbf{y}\}$, where E denotes expectation, and $\mathbf{y}$ the standardized vector deviate with variance matrix $E\{\mathbf{y}\mathbf{y}'\} = \mathbf{V}$. We have

$$E\{\mathbf{w}'\mathbf{y}\cdot\mathbf{h}'\mathbf{y}\} = E\{\mathbf{w}'\mathbf{y}\mathbf{y}'\mathbf{h}\} = \mathbf{w}'\mathbf{V}\mathbf{h} = \mathbf{w}'[(1-\rho) + \rho\mathbf{w}\mathbf{w}']\left[\frac{p\mathbf{d}}{\mathbf{w}'\mathbf{d}} - \mathbf{w}\right]$$

$$= p(1-\rho) + \rho p\mathbf{w}'\mathbf{w} - (1-\rho)\mathbf{w}'\mathbf{w} - \rho(\mathbf{w}'\mathbf{w})^2 = 0.$$

In view of this zero correlation the best discriminant function is of the form

$$(18) \qquad \frac{d_1}{v_1} y_1 + \frac{d_2}{v_2} y_2,$$

where $y_1 = \mathbf{w}'\mathbf{x}$ (the 'size' variable),

$y_2 = \mathbf{h}'\mathbf{x}$ (the 'shape' variable),

d_1 is the difference in means for y_1 and v_1 its variance, and similarly for y_2. Penrose has shown that even if $\mathbf{V}$ is not exactly of the homogeneous type (15), the above method often gives a very good discriminant function. Applying it to the numerical data referred to in section 3 above, for example, it will be found that we obtain estimates

	Size weighting ($d_1\mathbf{w}/v_1$)	Shape weighting ($d_2\mathbf{h}/v_2$)	Final weighting
x_1	1.4351	−2.3353	−0.9002
x_2	1.4351	−8.0664	−6.6313
x_3	1.4351	+5.9774	7.4125
x_4	1.4351	+4.4243	5.8594.

(19)

It should be noted that the final weightings in (19) correspond with formula (18), and differ slightly from those given by Penrose (Table 5), who makes allowance for the *observed* correlation between y_1 and y_2. This allowance seems

somewhat illogical and in any case rather a refinement. Thus Penrose's coefficients give a squared multiple correlation coefficient of 0.96334, whereas those in (19) give 0.96329 (compared with the maximum given in Section 3 of 0.96342).

This method is much quicker than the exact method, but of course the full analysis, as has been indicated in Section 3, enables the most efficient yet economical discriminant function to be found.

REFERENCES

[1] J. SHERMAN AND W. J. MORRISON, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix", *Annals of Math. Stat.*, Vol. 21 (1950), p. 124.

[2] J. SHERMAN AND W. J. MORRISON, "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix", abstract, *Annals of Math. Stat.*, Vol. 20 (1949), p. 621.

[3] M. S. BARTLETT, "Further aspects of the theory of multiple regression", *Proc. Camb. Phil. Soc.*, Vol. 34 (1938), p. 33.

[4] R. A. FISHER, "The use of multiple measurement in taxonomic problems", *Annals of Eugenics*, Vol. 7 (1936), p. 179.

[5] M. S. BARTLETT, "Multivariate analysis", *Jour. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), p. 176.

[6] S. PENROSE, "Some notes on discrimination", *Annals of Eugenics*, Vol. 13 (1946–47), p. 228.

NOTES

This section is devoted to brief research and expository articles and other short items.

———◆———

ON A THEOREM OF LYAPUNOV

By David Blackwell

Howard University

The purpose of this note is to point out two extensions of the following theorem of Lyapunov[1], and to note an interesting statistical consequence of each.[2]

LYAPUNOV'S THEOREM: *Let $u_1, \cdots, u_n$ be non-atomic[3] measures on a Borel field $\mathcal{B}$ of subsets of a space X. The set R of vectors $[u_1(E), \cdots, u_n(E)]$, $E \, \epsilon \, \mathcal{B}$, is convex, i.e., if $r_1, r_2 \, \epsilon \, R$, so does $tr_1 + (1 - t)r_2$ for $0 \leq t \leq 1$.*

EXTENSION 1. *Let $u_1, \cdots, u_n$ be non-atomic measures on a Borel field of subsets of a space X and let A be any subset of n-dimensional Euclidean space. Let $f = a(x) = [a_1(x), \cdots, a_n(x)]$ be any $\mathcal{B}$-measurable function defined on X with values in A, and define $v(f) = [\int a_1(x) \, du_1, \cdots, \int a_n(x) \, du_n]$. The set of vectors $v(f)$ is convex.*

Lyapunov's theorem is the special case in which A consists of two points $(0, \cdots, 0)$ and $(1, \cdots, 1)$.

PROOF. Let $v(f_i) = v_i$, $f_i = [a_{i1}(x), \cdots, a_{in}(x)]$, $i = 1, 2$, and consider the $2n$-dimensional measure

$$w(E) = \int_E a_{11}(x) \, du_1 \cdots, \int_E a_{1n}(x) \, du_n, \int_E a_{21}(x) \, du_1, \cdots, \int_E a_{2n}(x) \, du_n.$$

Since $w(N) = (0, \cdots, 0)$ where N is the null set, $w(X) = (v_1, v_2)$, for any t, $0 < t < 1$, there is, by Lyapunov's theorem, a set $E \, \epsilon \, \mathcal{B}$ with $w(E) = (tv_1, tv_2)$,

---

[1] "Sur les fonctions-vecteurs complètement additives," *Bull. Acad. Sci. URSS. Sér. Math.* Vol. 4 (1940), pp. 465–478. For a simplified proof of Lyapunov's results, see Halmos, "The range of a vector measure," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 416–421.

[2] Since this note was submitted, results obtained earlier by Dvoretzky, Wald, and Wolfowitz have appeared in the April 1950 *Proceedings of the National Academy of Sciences.* Their results are closely related to those presented here, and anticipate the general conclusion reached here: that in dealing with non-atomic distributions, mixed strategies are unnecessary. Their principal tool is also an extension of Lyapunov's theorem; their extension does not appear to contain or be contained in either of the extensions given here. The situation considered here is more general in that an infinite number of possible terminal actions are possible, but more restricted in that only mixtures of a finite number of pure strategies are considered here.

[3] A measure u is non-atomic if every set of non-zero measure has a subset of different non-zero measure.

so that $w(CE) = [(1 - t)v_1, (1 - t)v_2]$. Define $f = f_1$ on E, $f = f_2$ on CE. Then $v(f) = tv_1 + (1 - t)v_2$. This completes the proof.

This extension may be reformulated using statistical language, in the special case where $u_1, \cdots, u_n$ are probability measures, as follows: *In a statistical decision problem in which there are only a finite number of possible distributions, each of which is non-atomic, mixed strategies on the part of the statistician are unnecessary: anything which can be achieved with mixed strategies can already be achieved with pure strategies.*

In amplification, $u_1, \cdots, u_n$ are probability distributions, and x is an observation chosen according to one of them. Having observed x, the statistician must choose an action d from a set D of possible actions. His loss in choosing an action d is $a(1, d), \cdots, a(n, d)$ when the true distribution of x is $u_1, \cdots, u_n$, respectively. Thus the choice of d may be described as choosing a point $a \in A$, the subset of n-dimensional space consisting of the set of loss vectors

$$[a(1, d), \cdots, a(n, d)], d \in D.$$

Of course several points d may lead to the same a. From our point of view, two d's with the same a may be identified, so that it is no loss of generality to consider A itself as the set of possible actions.

A strategy for the statistician is then a function $f = a(x)$ from X into A, specifying the action to be taken (i.e., the loss vector to be chosen) when x is observed. We shall consider only $\mathcal{B}$-measurable strategies f. The expected loss vector from a strategy f is $v[f] = \int a_1(x) \, du_1, \cdots, \int a_n(x) \, du_n$; the i-th component is the expected loss from f when the true distribution is u_i. Thus the range R of $v(f)$ is the set of expected loss vectors attainable with pure strategies f. By mixed strategies, i.e., using strategies $f_1, \cdots, f_k$ with probabilities

$$p_1, \cdots, p_k, \ p_i \geq 1, \ \Sigma p_i = 1,$$

the statistician can attain all vectors in the convex set determined by R, and only those. Thus if R is already convex, nothing is gained by the use of mixed strategies.[4]

Sequential sampling. The above discussion applies directly only to the action to be taken after a sample point x has been obtained, sequentially or otherwise, and asserts that, in the non-atomic case, nothing is gained by mixing actions. It is still possible that a mixture of sampling plans, for instance tossing a coin to decide whether to take another observation, might, even with non-atomic distributions, achieve an expected loss vector not attainable with any one sampling plan. It turns out, however, that nothing is gained by mixing sampling plans, provided all sampling plans provide for at least one observation, and that the distributions of this observation are non-atomic. Formally, we have the

---

[4] It has been shown by the author in a paper submitted to the *Proceedings of the American Mathematical Society* that if A is closed, R is closed. Closure of R implies that a minimax strategy for the statistician exists.

THEOREM: *Let $x = (x_1, x_2, \cdots)$ be a sequence of chance variables whose joint distribution is one of n probability distributions $u_1, \cdots, u_n$. Let $S_1, \cdots, S_N$ be N sequential decision functions, each requiring the observation of x_1, and suppose the distributions of x_1 under $u_1, \cdots, u_n$ are non-atomic. Then any expected loss vector attainable from a mixture of $S_1, \cdots, S_N$ is also attainable from a single decision function S.*

PROOF. Let $d_{ij}(x)$ be the loss from S_j when the distribution of x is u_i. (The loss is a function of x as well as i, j, since the cost of observations may vary with x.) Then $a_j = (Ed_{ij}, \cdots, Ed_{nj})$ is the expected loss vector from S_j. Since $S_1, \cdots, S_N$ all involve observing x_1, the statistician need not make up his mind about which decision procedure to use until after x_1 is observed, i.e., a possible decision procedure is a division $\mathfrak{D}$ of sample space into N mutually exclusive x_1-sets $D_1, \cdots, D_N$, and to use decision procedure S_j if $x_1 \in D_j$. The expected loss vector from $\mathfrak{D}$ is

$$v(\mathfrak{D}) = \left(\sum_{j=1}^{N} \int_{D_j} \phi_{1j}(x_1) \, du_1(x_1), \cdots, \sum_{j=1}^{N} \int_{D_j} \phi_{nj}(x_1) \, du_n(x_1) \right),$$

where $\phi_{ij}(x_1)$ is the conditional expectation of d_{ij} with respect to x_1. If $\mathfrak{D}$ is the decision procedure with $D_j =$ space X, $D_i =$ null set for $i \neq j$, then $v(\mathfrak{D}) = a_j$. Thus it is sufficient to show that the range of $v(\mathfrak{D})$ is convex.

The convexity of the range of $v(\mathfrak{D})$ is the special case where $u_1, \cdots, u_n$ are probability measures of

EXTENSION 2. *Let $u_1, \cdots, u_n$ be non-atomic measures on a Borel field $\mathfrak{B}$ of subsets of a space X, let $\phi_{ij}(x)$, $i = 1, \cdots, n$, $j = 1, \cdots, N$, be $\mathfrak{B}$-measurable functions of x such that ϕ_{ij} is u_i-integrable over X, let $\mathfrak{D} = (D_1, \cdots, D_N)$ be a decompositon of X into N disjoint subsets, and define*

$$v(\mathfrak{D}) = \left(\sum_{j=1}^{N} \int_{D_j} \phi_{1j} \, du_1, \cdots, \sum_{j=1}^{N} \int_{D_j} \phi_{nj} \, du_n \right).$$

The range of $v(\mathfrak{D})$ is convex.

PROOF. Let $\mathfrak{D}_k = (D_{k1}, \cdots, D_{kN})$, $k = 1, 2$ be two decompositions. We must show that for any t, $0 \leq t \leq 1$, there is a $\mathfrak{D}$ with $v(\mathfrak{D}) = tv(\mathfrak{D}_1) + (1 - t)v(\mathfrak{D}_2)$. Write $m_{ij}(B) = \int_B \phi_{ij} \, du_i$, and consider the $2nN$-dimensional measure $w(B) = m_{ij}(BD_{kj})$, $i = 1, \cdots, n$, $j = 1, \cdots, N$, $k = 1, 2$. Since $w(B)$ is non-atomic, Lyapunov's theorem asserts there is a B with $w(B) = tw(x)$, i.e., $m_{ij}(BD_{kj}) = tm_{ij}(D_{kj})$. Then $m_{ij}(C(B)D_{kj}) = (1 - t)m_{ij}(D_{kj})$. Define $D_j = BD_{1j} + C(B)D_{2j}$, $j = 1, \cdots, N$, $\mathfrak{D} = (D_1, \cdots, D_N)$. Then

$$v(\mathfrak{D}) = \sum_{j=1}^{N} [m_{1j}(D_j), \cdots, m_{nj}(D_j)]$$

$$= t \sum_{j=1}^{N} [m_{1j}(D_{1j}), \cdots, m_{nj}(D_{1j})] + (1 - t) \sum_{j=1}^{N} [m_{1j}(D_{2j}), \cdots, m_{nj}(D_{2j})]$$

$$= tv(\mathfrak{D}_1) + (1 - t)v(\mathfrak{D}_2).$$

A NOTE ON THE TEST OF SERIAL CORRELATION COEFFICIENTS

By Masami Ogawara

Meteorological-Research Institute, Tokyo

1. Summary. In this note the author points out that in the case of stationary Gaussian Markov process, i.e., autoregressive stochastic process, we can test the serial correlation coefficients by a method based on normal regression theory. Particularly, in the case of simple Markov process, we can find the confidence limits for its autocorrelation coefficient.

In this method, so far as random variables are concerned, not all the information in the original data is used, with a consequent reduction of degrees of freedom. However, the other part of information is introduced as parameters in the distribution functions of random variables and in the statistic useful for tests.

2. Introduction. For the test of the serial correlation coefficient, a method based on its distribution may be orthodox. Up to the present, however, many investigations along this line, e.g. R. L. Anderson [1], M. H. Quenouille [2], P. A. P. Moran [3], T. W. Anderson [4] and others seem to be confined in at least one of the following restrictions:

(1) circular definition,

(2) significance test, i.e., testing the uncorrelatedness of the process,

(3) approximate distribution.

In this paper, we do not use the distribution of a serial correlation coefficient itself, but normal regression theory, and will give the general testing method for an autoregressive stochastic process.

3. Fundamental theorems. The following theorems are fundamental in our method.

THEOREM 1. *Let* $x_t(t = \cdots, -1, 0, 1, 2, \cdots)$ *be a simple Markov process. If the values of* $x_{2k-1}(k = 1, 2, \cdots, n + 1)$ *are fixed, the random variables* x_{2k} *($k = 1, 2, \cdots, n$) are mutually independent.*[1]

This theorem is easily proved from the following facts:

(1) When the value of x_0 is given, $x_1, \cdots, x_n$ are independent of $x_{-1}, x_{-2}, \cdots$.

(2) When x_0 is given, the stochastic sequence $x_1, x_2, \cdots$, is also a simple Markov process for the inversely directed time scale.

Similarly, the following general theorem holds:

THEOREM 2. *Let* x_t *($t = \cdots, -1, 0, 1, 2, \cdots$) be a Markov process of order h. Then, if the values of* $x_{k(h+1)-h}, \cdots, x_{k(h+1)-1}, x_{k(h+1)+1}, \cdots, x_{k(h+1)+h}$ *($k = 1, 2, \cdots, n$) are given, the random variables* $x_{k(h+1)}$ *($k = 1, 2, \cdots, n$) are mutually independent.*

---

[1] This fact has been used by U. V. Linnik (without proof) in his proof of the central limit theorem for simple Markov process. *Izvestiya Akad. Nauk. USSR., Ser. Mat.*, Vol. 13 (1949).

THEOREM 3.[2] *Let x_t ($t = \cdots, -1, 0, 1, \cdots$) be a stationary Gaussian process. A necessary and sufficient condition that x_t should be a non-singular Markov process of order h is that its autocorrelation coefficients ρ_r satisfy the finite difference equation*

$$(1) \qquad \rho_r + a_1 \rho_{r-1} + \cdots + a_h \rho_{r-h} = 0, \qquad\qquad r = 1, 2, \cdots; a_h \gtrless 0,$$

where the a's are such that every root of the equation

$$z^h + a_1 z^{h-1} + \cdots + a_{h-1} z + a_h = 0$$

lies within the unit circle.

4. The case of a stationary Gaussian simple Markov process. Let m, σ^2 and $\rho_r (\equiv \rho^r)$ be the mean, variance and autocorrelation coefficient, respectively, of a stationary Gaussian simple Markov process x_t with discrete parameter t. According to Theorem 1, when the values of $x_{2k-1}(k = 1, 2, \cdots, n + 1)$ are fixed, $x_{2k}(k = 1, 2, \cdots, n)$ are mutually independent and, in this case, their conditional probability densities are given by

$$f(x_{2k} \mid x_{2k-1}, x_{2k+1}) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2} \{x_{2k} - (a + bx_k')\}^2 \right]$$

$$(k = 1, 2, \cdots, n),$$

where

$$(2) \qquad \begin{aligned} a &= m(1 - \rho)^2/(1 + \rho^2), \\ b &= 2\rho/(1 + \rho^2), \\ \sigma_0^2 &= \sigma^2(1 - \rho^2)/(1 + \rho^2), \\ x_k' &= (x_{2k-1} + x_{2k+1})/2. \end{aligned}$$

Considering x_k' as the fixed variates and applying normal regression theory [6], the maximum likelihood estimates of the parameter a, b, and σ_0^2 are given by

$$(3) \qquad \begin{aligned} \hat{a} &= \bar{x}_2 - \hat{b}\bar{x}', \\ \hat{b} &= \sum_{k=1}^n (x_k' - \bar{x}')(x_{2k} - \bar{x}_2) \Big/ \sum_{k=1}^n (x_k' - \bar{x}')^2, \\ \hat{\sigma}_0^2 &= \sum_{k=1}^n (x_{2k} - \hat{a} - \hat{b}x_k')^2/n, \end{aligned}$$

where

$$\bar{x}_2 = \sum_1^n x_{2k}/n, \quad \bar{x}' = \sum_1^n x_k'/n = (\bar{x}_1 + \bar{x}_3)/2$$

---

[2] M. Ogawara [5].

with

$$\bar{x}_1 = \sum_1^n x_{2k-1}/n, \quad \bar{x}_3 = \sum_1^n x_{2k+1}/n.$$

We can rewrite $\hat{b}$ as follows:

(4)
$$\hat{b} = 2r_1/(1 + r_2),$$

where

(5)
$$r_1 = \frac{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)(x_{2k} - \bar{x}_2) + \frac{1}{n}\sum_1^n (x_{2k} - \bar{x}_2)(x_{2k+1} - \bar{x}_3)\right\}}{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)^2 + \frac{1}{n}\sum_1^n (x_{2k+1} - \bar{x}_3)^2\right\}},$$

$$r_2 = \frac{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)(x_{2k+1} - \bar{x}_3)}{\frac{1}{2}\left\{\frac{1}{n}\sum_1^n (x_{2k-1} - \bar{x}_1)^2 + \frac{1}{n}\sum_1^n (x_{2k+1} - \bar{x}_3)^2\right\}}.$$

Because

$$\frac{\partial(a, b, \sigma_0^2)}{\partial(m, \sigma^2, \rho)} = \frac{2(1 - \rho)^2(1 - \rho^2)^2}{(1 + \rho^2)^4} > 0 \qquad (\text{for } |\rho| \lessgtr 1),$$

the maximum likelihood estimates of m, σ^2 and ρ are given by

(6)
$$\hat{m} = \hat{a}/(1 - \hat{b}),$$
$$\hat{\sigma}^2 = \hat{\sigma}_0^2/\sqrt{1 - \hat{b}^2},$$
$$\hat{\rho} = (1 - \sqrt{1 - \hat{b}^2})/\hat{b}.$$

Since, as the function of random variables x_{2k},

(7)
$$F = \frac{(\hat{b} - b)^2 \sum_1^n (x_k' - \bar{x}')^2 \cdot (n - 2)}{\sum_1^n (x_{2k} - \hat{a} - \hat{b}x_k')^2}$$

has the F-distribution with 1 and $n - 2$ degrees of freedom, we can test the hypotheses $\rho = \rho_0$ or $b = b_0 = 2\rho_0/(1 + \rho_0^2)$. As the function $\rho = (1 - \sqrt{1 - b^2})/b$ is monotone increasing, we can also find confidence limits for ρ from those for b.

5. The case of a stationary Gaussian Markov process of order h. Let, as before, m, σ^2 and ρ_r be the mean, variance and autocorrelation coefficient of our process x_t respectively. From Theorem 2, the random variables $x_{k(h+1)}(k = 1, 2, \cdots, n)$ are independent of each other, under the condition that the variables $x_{k(h+1)-p}$, $x_{k(h+1)+p}$ $(p = 1, 2, \cdots, h; k = 1, 2, \cdots, n)$ are fixed, and, in the present case, their conditional probability densities are given by

$$f(x_{k(h+1)} \mid x_{k(h+1)-p}, x_{k(h+1)+p} ; p = 1, 2, \cdots, h)$$

(8)
$$= \frac{1}{\sqrt{2\pi}\,\sigma_0^2} \exp\left[-\frac{1}{2\sigma_0^2}\left\{x_{k(h+1)-p} - \sum_{p=0}^h b_p x_{pk}'\right\}^2\right] \qquad (k = 1, 2, \cdots, n),$$

where $x'_{pk} = (x_{k(h+1)-p} + x_{k(h+1)+p})/2$ $(p = 1, 2, \cdots, h)$, $x'_{0k} = 1$, and where

$$b_0 = m\left(1 - 2\sum_{p=1}^{h} c_p\right), \qquad b_p = 2c_p \ (p = 1, 2, \cdots, h),$$

(9)
$$
\begin{pmatrix} c_h \\ c_{h-1} \\ \vdots \\ c_1 \\ c_1 \\ \vdots \\ c_h \end{pmatrix}
=
\begin{pmatrix}
1 & \cdots & \rho_{h-1} & \rho_{h+1} & \cdots & \rho_{2h} \\
\ddots & & \vdots & \vdots & & \vdots \\
\rho_{h-1} & \cdots & 1 & \rho_2 & \cdots & \rho_{h+1} \\
\rho_{h+1} & \cdots & \rho_2 & 1 & \cdots & \rho_{h-1} \\
\vdots & & \vdots & \vdots & & \vdots \\
\rho_{2h} & \cdots & \rho_{h+1} & \rho_{h-1} & \cdots & 1
\end{pmatrix}^{-1}
\cdot
\begin{pmatrix} \rho_h \\ \rho_{h-1} \\ \vdots \\ \rho_1 \\ \rho_1 \\ \vdots \\ \rho_h \end{pmatrix}
$$

and

(10)
$$\sigma_0^2 = \frac{1 + a_1\rho_1 + \cdots + a_h\rho_h}{1 + a_1^2 + \cdots + a_h^2}\,\sigma^2,$$

where the a's are the coefficients of equation (1).

Considering the relations (1) and (9), the hypotheses concerning $\rho_1, \cdots, \rho_h$[3] is equivalent to the hypotheses concerning $c_1, \cdots, c_h$ or $b_1, \cdots, b_h$. Thus normal regression theory is applicable.

Moreover, we can estimate the order of the Markov process as follows. The above stated theory holds whenever the essential order h_0 of the process is not greater than h. Hence, we may select as its order such a value h_0 that the hypotheses $b_{h_0} = b_{h_0+1} = \cdots = b_h = 0$ is rejected but the hypotheses $b_{h_0+1} = b_{h_0+2} = \cdots = b_h = 0$ is not rejected.

REFERENCES

[1] R. L. ANDERSON, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13.

[2] M. H. QUENOUILLE, "Some results in the testing of serial correlation coefficients," *Biometrika*, Vol. 35 (1948), pp. 261–267.

[3] P. A. P. MORAN, "Some theorems on time series, II. The significance of the serial correlation coefficient," *Biometrika*, Vol. 35 (1948), pp. 255–260.

[4] T. W. ANDERSON, "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, Vol. 31 (1948), pp. 88–116.

[5] M. OGAWARA, "On the normal stationary Markov process of higher order," *Bull. of Math. Stat.* (in Japanese), Vol. 2 (1948), pp. 42–46.

[6] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943.

[3] Owing to (1), this is also equivalent to the hypotheses concerning $a_1, \cdots, a_h$.

REMARK ON SEPARABLE SPACES OF PROBABILITY MEASURES

By Agnes Berger

New York City

Early writers on mathematical statistics often had to assume that the distributions under consideration either admitted probability densities, sometimes subject to further regularity conditions, or that they were purely discrete; in general, two separate arguments were needed to deal with the two cases. More recent authors however have achieved greater generality and, at the same time, a unification of methods by dispensing altogether with assumptions on the distributions themselves and specifying, instead, their relation to each other. In particular, these writers assume (for example in [1], [2], [3]) that the probability measures under consideration form what is sometimes called a "dominated set of measures", defined as follows: Let X be the sample space, $\mathfrak{B}$ a Borel field of some subsets of X and let $\Omega = \{m\}$ be a set of probability measures defined on $\mathfrak{B}$. Ω is called a dominated set of measures if there exists a σ-finite measure μ such that every m in Ω is absolutely continuous with respect to μ.

One of the important consequences of assuming that Ω be dominated is that, if such an Ω is metrized by introducing

$$d(m, m') = \sup_{B \cdot \mathfrak{B}} | m(B) - m'(B) |$$

as a metric and $\mathfrak{B}$ is a separable Borel field (as for instance in the case of Borel sets in finite dimensional Euclidean spaces), then Ω is separable with respect to the topology induced by d. (Proof of this can be based on Hopf's approximation theorem as indicated in [1]; a proof for measures dominated by Lebesgue measure is referred to at the end of [4].)

Since the separability of dominated sets of measures is used to great advantage (for example in [1] and in [4]), one wonders whether there exist any other separable sets of measures than dominated ones. It will be shown to the contrary, that an even weaker separability condition than the one described implies that the set be dominated. In order to state the exact theorem, we shall consider a set $M = \{m\}$ of probability-measures defined on a common Borel field $\mathfrak{B}$ of subsets of some abstract space X and introduce a weak topology into M in the usual way (see [5]) by defining a complete system of neighborhoods as follows: For every p in M and for every finite collection of sets $B_1, B_2, \cdots, B_k$ in $\mathfrak{B}$ and every $\epsilon > 0$, let $\alpha = (B_1, B_2, \cdots, B_k; \epsilon)$ and let

$$V_\alpha(p) = \{m \text{ in } M || m(B_i) - p(B_i) | < \epsilon, i = 1, 2, \cdots, k,\},$$

i.e. the set of all those measures in M whose values assumed on the sets B_1, $B_2, \cdots, B_k$ differ less than ϵ in absolute value from the corresponding values of p. $V_\alpha(p)$ is called the neighborhood of index α of the measure p. By letting α range over all possible finite collection of sets in $\mathfrak{B}$ and all positive numbers ϵ, $V_\alpha(p)$ defines a complete system of neighborhoods (see for instance [6]), so that M may be regarded as a topological space. We shall prove the following theorem:

THEOREM. *If a set of measures M is separable with respect to the weak topology defined above then M is dominated.*

PROOF. By assumption, there exists a sequence of measures $\{m_i\}$ in M such that to any given p in M and any given α, there exists an m_i in $V_\alpha(p)$. Let $\mu = \sum_{i=1}^{\infty} c_i m_i$, $0 < c_i < 1$, $\sum_{i=1}^{\infty} c_i = 1$; then $\mu(X) = 1$. Let B in $\mathfrak{B}$ be such that $\mu(B) = 0$. Obviously, $m_i(B) = 0$ for all i. Let p be an arbitrary fixed measure in M and consider the sequence of neighborhoods $V_{\alpha_j}(p)$, where $\alpha_j = \left(B; \dfrac{1}{2^j}\right)$ $j = 1, 2, \cdots$. Then for any fixed j there exists an m_k which is in $V_{\alpha_j}(p)$, thus

$$| m_k(B) - p(B) | < \frac{1}{2^j}.$$

Since $m_i(B) = 0$ for all i, $p(B) < \dfrac{1}{2^j}$ and since j was arbitrary this means $p(B) = 0$. Thus whenever $\mu(B) = 0$ for some B we have $p(B) = 0$ for every p in M, as we wanted to prove.

Since a set of measures separable with respect to the metric topology induced by

$$d(m, m') = \sup_{B \cdot \mathfrak{B}} | m(B) - m'(B) |$$

is a fortiori separable in the weak topology, we can add the following theorem:

THEOREM. *A necessary and sufficient condition for a set of measures defined over a separable Borel field to be separable with respect to the topology induced by the metric d is that it be a dominated set of measures.*

REFERENCES

[1] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation," *Sankhyā*, to be published.
[2] P. R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), p. 225.
[3] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), p. 1.
[4] ABRAHAM WALD, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), p. 165.
[5] J. VON NEUMANN, "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren," *Mathematische Annalen*, Vol. 102 (1930), p. 370.
[6] HAUSDORFF, *Grundzüge der Mengenlehre*, Leipzig, 1914.
[7] E. HOPF, "Ergodentheorie," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Vol. 5, No. 2 (1937), p. 4.

TABLE OF THE ASYMPTOTIC DISTRIBUTION
OF THE SECOND EXTREME

By E. J. Gumbel and J. Arthur Greenwood

New York City and Manhattan Life Insurance Company

The asymptotic distributions of the extreme values taken from an initial distribution of the exponential type are now widely used, for example in flood control [6] and in problems connected with the breaking strength of material [1]. Therefore, the corresponding distribution of the penultimate (and of the second) value may also be of practical interest.

Let $F(x)$ be the initial probability; let $f(x) = F'(x)$ be the initial density (distribution). Let n be a large sample size; let the rank $m(m << n)$ be counted from the top. Finally, let the parameters u_m and α_m be defined as the solutions of

$$(1) \qquad F(u_m) = 1 - m/n; \qquad \alpha_m = nf(u_m)/m.$$

Then the asymptotic distribution $\varphi_m(x_m)$ of the mth largest value x_m is [2]

$$\varphi_m(x_m) = \frac{m^m}{\Gamma(m)} \alpha_m \exp[-my_m - me^{-y_m}],$$

where

$$y_m = \alpha_m(x_m - u_m),$$

provided that the initial distribution is of the exponential type. The asymptotic distribution $_m\varphi(_m x)$ of the mth smallest value is

$$_m\varphi(_m x) = \varphi_m(-x_m).$$

The probability function $\Phi_m(x_m)$ is obtained from

$$\Phi_m(x_m) = \frac{m^m}{\Gamma(m)} \int_{-\infty}^{y_m} \exp[-my - me^{-y}] \, dy$$

$$= \frac{1}{\Gamma(m)} \int_{me^{-y_m}}^{\infty} z^{m-1} e^{-z} \, dz,$$

whence

$$(2) \qquad \Phi_m(x_m) = 1 - I(t_m, m - 1),$$

where

$$t_m = \sqrt{m}\, e^{-y_m}$$

and I is the incomplete Gamma function ratio of Karl Pearson [5]. In the special case $m = 2$, the probability function of the penultimate value is

$$(3) \qquad \Phi_2(x_2) = 1 - I(\sqrt{2}\, e^{-y_2}, 1).$$

The modal penultimate value is, of course, u_2, and the intervals corresponding

TABLE I
Probability $\Phi_2(y_2)$ of the penultimate value y_1

y_1	Φ_1	β^2	y_1	Φ_1	β^2	y_1	Φ_1	β^2
−1.95	.00001		0.55	.67935	− 89	3.05	.99579	− 4
−1.90	.00002		0.60	.69990	91	3.10	.99618	− 4
−1.85	.00004		0.65	.71954	91	3.15	.99653	
−1.80	.00007		0.70	.73827	92	3.20	.99685	
−1.75	.00013		0.75	.75608	− 92	3.25	.99714	
−1.70	.00021	+ 5	0.80	.77297	90	3.30	.99741	
−1.65	.00034	7	0.85	.78896	89	3.35	.99765	
−1.60	.00054	10	0.90	.80406	87	3.40	.99787	
−1.55	.00084	14	0.95	.81829	85	3.45	.99807	
−1.50	.00128	+ 17	1.00	.83167	− 81	3.50	.99825	
−1.45	.00189	24	1.05	.84424	80	3.55	.99841	
−1.40	.00274	30	1.10	.85601	75	3.60	.99856	
−1.35	.00389	38	1.15	.86703	74	3.65	.99869	
−1.30	.00542	47	1.20	.87731	69	3.70	.99882	
−1.25	.00742	+ 56	1.25	.88690	− 65	3.75	.99893	
−1.20	.00998	68	1.30	.89584	64	3.80	.99903	
−1.15	.01322	77	1.35	.90414	59	3.85	.99912	
−1.10	.01723	89	1.40	.91185	55	3.90	.99920	
−1.05	.02213	100	1.45	.91901	54	3.95	.99928	
−1.00	.02803	+ 110	1.50	.92563	− 49	4.00	.99935	
−0.95	.03503	121	1.55	.93176	46	4.05	.99941	
−0.90	.04324	129	1.60	.93743	44	4.10	.99946	
−0.85	.05274	135	1.65	.94266	40	4.15	.99951	
−0.80	.06359	142	1.70	.94749	39	4.20	.99956	
−0.75	.07586	+ 145	1.75	.95193	− 34	4.25	.99960	
−0.70	.08958	147	1.80	.95603	34	4.30	.99964	
−0.65	.10477	145	1.85	.95979	29	4.35	.99967	
−0.60	.12141	142	1.90	.96326	29	4.40	.99970	
−0.55	.13947	138	1.95	.96644	27	4.45	.99973	
−0.50	.15891	+ 130	2.00	.96935	− 23	4.50	.99976	
−0.45	.17965	121	2.05	.97203	23	4.55	.99978	
−0.40	.20160	111	2.10	.97448	20	4.60	.99980	
−0.35	.22466	99	2.15	.97673	19	4.65	.99982	
−0.30	.24871	86	2.20	.97879	18	4.70	.99984	
−0.25	.27362	+ 71	2.25	.98067	− 16	4.75	.99985	
−0.20	.29924	57	2.30	.98239	14	4.80	.99987	
−0.15	.32543	44	2.35	.98397	15	4.85	.99988	
−0.10	.35206	29	2.40	.98540	12	4.90	.99989	
−0.05	.37898	11	2.45	.98671	11	4.95	.99990	
0	.40601	+ 1	2.50	.98791	− 11	5.00	.99991	
0.05	.43305	− 13	2.55	.98900	9	5.05	.99992	
0.10	.45996	25	2.60	.99000	9	5.10	.99993	
0.15	.48662	37	2.65	.99091	8	5.15	.99993	
0.20	.51291	47	2.70	.99174	8	5.20	.99994	
0.25	.53873	− 56	2.75	.99249	− 6	5.25	.99995	
0.30	.56399	65	2.80	.99318	7	5.35	.99996	
0.35	.58860	72	2.85	.99380	5	5.50	.99997	
0.40	.61249	77	2.90	.99437	5	5.65	.99998	
0.45	.63561	82	2.95	.99489	5	5.90	.99999	
0.50	.65791	− 86	3.00	.99536	− 4	6.45	1.00000	

TABLE II
Probability points

$\Phi_2(y_2)$	y_2	$\Phi_2(y_2)$	y_2
.005	−1.31239	.995	2.96138
.010	−1.19972	.990	2.59995
.025	−1.02454	.975	2.11110
.050	−0.86371	.950	1.72777
.100	−0.66519	.900	1.32461
.250	−0.29737	.750	0.73264
.500	0.17534		

to the probabilities $P_1 = 0.68269$, $P_2 = 0.95445$, and $P_3 = .99730$ are $y_2 = \pm 0.75409$, $y_2' = 1.78196$, $y_2'' = 3.27883$, respectively.

The present five-decimal table was computed by interpolation in Pearson's table. The last six lines indicate the first values of y_2 for which Φ_2 differs from the value indicated by less than $5 \cdot 10^{-6}$. The table was checked by differencing and by comparison with the short table of percentage points (Table II) which was

SCHEME OF A PROBABILITY PAPER FOR THE PENULTIMATE AND SECOND VALUE.
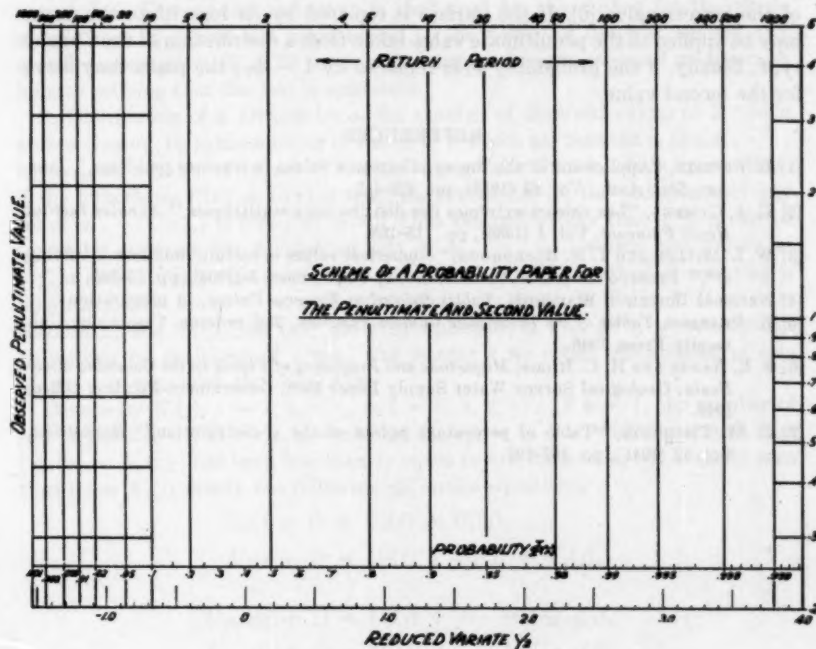
FIGURE 1

computed by noting that

$$2me^{-y_m}$$

has the χ^2 distribution with $2m$ degrees of freedom, and so transforming the percentage points given by Thompson [7] (pp. 188–189, line $y = 4$), setting $y_2 = \ln 4 - \ln \chi^2$.

More decimal places may be obtained by direct substitution in (3), by use of the relation

$$(4) \qquad\qquad \Phi_2(x_2) = \Phi_1(z) + \varphi_1(z),$$

where $z = y_2 - \ln 2$ and Φ_1 and φ_1, respectively, the probability and density of the largest value, are given in a seven-decimal table originally calculated by Greenwood [4], and from the nine-decimal table of $(x + 1)e^{-x}$ by Miller and Rosebrugh [3], pp. 80–101, where

$$x = 2e^{-y_2}.$$

Table I is basic for the construction of a probability paper (Figure 1) for the penultimate value which can be used in the same way as the probability paper of the largest value [6]. If the variate is replaced by its logarithm, the paper may be applied to the penultimate value taken from a distribution of the Cauchy type. Finally, if the probability Φ_2 is replaced by $1 - \Phi_2$, the paper may serve for the second value.

REFERENCES

[1] B. EPSTEIN, "Application of the theory of extreme values in fracture problems," *Jour. Am. Stat. Assn.*, Vol. 43 (1948), pp. 403–412.

[2] E. J. GUMBEL, "Les valeurs extrêmes des distributions statistiques," *Annales Institut Henri Poincaré*, Vol. 4 (1935), pp. 115–158.

[3] W. L. MILLER AND T. R. ROSEBRUGH, "Numerical values of certain functions involving e^{-x}," *Trans. of Roy. Soc. Canada*, Ser. 2, Vol. 9, Sect. 3 (1903), pp. 73–107.

[4] National Bureau of Standards, *Tables Related to Extreme Values*, in preparation.

[5] K. PEARSON, *Tables of the Incomplete Gamma Function*, 2nd re-issue, Cambridge University Press, 1946.

[6] S. E. RANTZ AND H. C. RIGGS, *Magnitude and Frequency of Floods in the Columbia River Basin*, Geological Survey Water Supply Paper 1080, Government Printing Office, 1949.

[7] C. M. THOMPSON, "Table of percentage points of the χ^2-distribution," *Biometrika*, Vol. 32 (1941), pp. 187–191.

THE DISTRIBUTION OF THE MAXIMUM DEVIATION BETWEEN TWO SAMPLE CUMULATIVE STEP FUNCTIONS

By Frank J. Massey, Jr.[1]

University of Oregon

1. Summary. Let $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_m$ be the ordered results of two random samples from populations having continuous cumulative distribution functions $F(x)$ and $G(x)$ respectively. Let $S_n(x) = k/n$ when k is the number of observed values of X which are less than or equal to x, and similarly let $S'_m(y) = j/m$ where j is the number of observed values of Y which are less than or equal to y.

The statistic $d = \max_x | S_n(x) - S'_m(x) |$ can be used to test the hypothesis $F(x) \equiv G(x)$, where the hypothesis would be rejected if the observed d is significantly large. The limiting distribution of $d \sqrt{\dfrac{mn}{m+n}}$ has been derived [1] and [4], and tabled [5]. In this paper a method of obtaining the exact distribution of d for small samples is described, and a short table for equal size samples is included. The general technique is that used by the author for the single sample case [2]. There is a lower bound to the power of the test against any specified alternative, [3]. This lower bound approaches one as n and m approach infinity proving that the test is consistent.

2. Distribution of d. Denote by α_1 the number of observed values of Y which are less than x_1, by α_2 the number of values of Y which are between x_1 and x_2, $\cdots$, by α_{n+1} the number of values of Y which are greater than x_n. It is known that, if the hypothesis $F(x) \equiv G(x)$ is true, the probability of the occurrence of any set of $\alpha_1, \cdots, \alpha_{n+1}$ is $n!m!/(m+n)!$ Thus the probability that $d \leq a$ can be found by counting the number of sets of $\alpha_1, \cdots, \alpha_{n+1}$ which give values of $d \leq a$ and multiply this number by $n!m!/(m+n)!$ The method of counting is illustrated here for $n = m$, and some results are given in Table 1. If $n = m$ then $S_n(x)$ and $S'_m(y)$ can only differ by multiples of $1/n$. (If $n \neq m$ they can only differ by multiples of $1/mn$.) For integer k we count the number of sets of $\alpha_1, \cdots, \alpha_{n+1}$ such that $d \leq k/n$.

Denote by $U_i(j), j = 1, 2, \cdots, n, i = 0, 1, 2, \cdots, 2k-1$, the number of sets of possible $\alpha_1, \alpha_2, \cdots, \alpha_j$ such that $S'_n(x_j) = (j + i - k)/n$ and such that $| S_n(x) - S'_n(x) |$ has been less than or equal to k/n for $x < x_j$. It is easily seen that these $X_i(j)$ satisfy the following difference equations.

$$U_0(j + 1) = U_0(j) + U_1(j),$$
$$U_1(j + 1) = U_0(j) + U_1(j) + U_2(j),$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$U_{2k-2}(j + 1) = U_0(j) + \cdots + U_{2k-1}(j),$$
$$U_{2k-1}(j + 1) = U_0(j) + \cdots + U_{2k-1}(j).$$

TABLE 1

Probability of $d \leq k/n$

$n = m$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
1	1.000000					
2	.666667	1.000000				
3	.400000	.900000	1.000000			
4	.228571	.771429	.971429	1.000000		
5	.126984	.642857	.920635	.992063	1.000000	
6	.069264	.525974	.857143	.974026	.997835	1.000000
7	.037296	.424825	.787879	.946970	.991841	.999417
8	.019891	.339860	.717327	.912976	.981352	.997514
9	.010537	.269889	.648293	.874126	.966434	.993706
10	.005542	.213070	.582476	.832179	.947552	.987659
11	.002903	.167412	.520850	.788524	.925339	.979261
12	.001515	.131018	.463902	.744225	.900453	.968564
13	.000788	.102194	.411804	.700080	.873512	.955728
14	.000408	.079484	.364515	.656680	.845065	.940970
15	.000211	.061669	.321862	.614453	.815584	.924536
16	.000109	.047744	.283588	.573707	.785465	.906674
17	.000056	.036893	.249393	.534647	.755041	.887623
18	.000029	.028460	.218952	.497410	.724582	.867606
19	$.0^4148$	.021922	.191938	.462071	.694311	.846827
20	$.0^5761$	.016863	.168030	.428664	.664409	.825467
21	$.0^5390$	.012956	.146921	.397187	.635020	.803688
22	$.0^5199$	.009943	.128321	.367614	.606260	.781632
23	$.0^5102$	.007623	.111963	.339899	.578218	.759422
24	$.0^652$	.005839	.097600	.313983	.550963	.737166
25	$.0^627$	.004468	.085007	.289796	.524546	.714958
26	$.0^614$	.003417	.073980	.267263	.499005	.692877
27	$.0^769$	.002611	.064338	.246303	.474362	.670992
28	$.0^735$	.001994	.055914	.226833	.450633	.649362
29	$.0^718$	.001522	.048563	.208772	.427823	.628036
30	$.0^891$	.001161	.042154	.192037	.405929	.607055
31	$.0^846$	.000885	.036570	.176546	.384946	.586455
32	$.0^823$	.000674	.031710	.162223	.364861	.566264
33	$.0^812$	.000513	.027483	.148989	.345657	.546505
34	$.0^960$	.000391	.023808	.136773	.327316	.527198
35	$.0^931$	.000297	.020616	.125505	.309816	.508355
36	$.0^916$	.000226	.017845	.115120	.293133	.489989
37	$.0^{10}79$	.000172	.015440	.105553	.277243	.472107
38	$.0^{10}40$	.000131	.013355	.096747	.262121	.454713
39	$.0^{10}20$	.000099	.011547	.088645	.247738	.437811
40	$.0^{10}10$	.000075	.009981	.081195	.234069	.421400

TABLE 1—*Continued*

n = m	k = 7	k = 8	k = 9	k = 10	k = 11	k = 12
1						
2						
3						
4						
5						
6						
7	1.000000					
8	.999845	1.000000				
9	.999260	.999959	1.000000			
10	.997943	.999783	.999989	1.000000		
11	.995634	.999345	.999938	.999997	1.000000	
12	.992141	.998503	.999796	.999982	.999999	1.000000
13	.987351	.997125	.999500	.999938	.999995	1.000000
14	.981218	.995100	.998979	.999837	.999981	.999999
15	.973752	.992344	.998163	.999647	.999948	.999994
16	.965002	.988801	.996985	.999330	.999880	.999983
17	.955047	.984439	.995389	.998847	.999762	.999960
18	.943982	.979252	.993331	.998160	.999571	.999917
19	.931911	.973251	.990776	.997233	.999286	.999844
20	.918942	.966458	.987701	.996033	.998884	.999729
21	.905183	.958911	.984095	.99453	.99834	.99956
22	.890738	.950653	.979953	.99271	.99764	.99933
23	.875705	.941731	.975280	.99055	.99676	.99901
24	.860177	.932197	.970087	.98803	.99568	.99860
25	.844240	.922101	.964389	.98516	.99438	.99808
26	.827971	.911498	.958206	.98193	.99287	.99744
27	.811443	.900437	.951562	.97833	.99111	.99667
28	.794722	.888969	.944481	.97438	.98911	.99576
29	.777865	.877140	.936989	.97007	.98686	.99469
30	.760927	.864996	.929113	.96542	.98436	.99346
31	.743955	.852580	.920880	.96044	.98160	.9921
32	.726992	.839930	.912319	.95514	.97859	.9905
33	.710076	.827086	.903455	.94953	.97533	.9888
34	.693242	.814080	.894315	.94363	.97182	.9868
35	.676519	.800946	.884924	.93745	.96807	.9847
36	.659934	.787713	.875307	.93101	.96407	.9824
37	.643512	.774409	.865487	.92432	.95985	.9799
38	.627273	.761059	.855487	.91740	.95540	.9773
39	.611234	.747687	.845327	.91027	.95074	.9744
40	.595413	.734313	.835029	.90293	.94587	.9714

For small n these equations can be solved by iteration, which was done in constructing Table 1. Initial conditions an $U_k(0) = 1$, $U_i(0) = 0$ for $i \neq k$. It might be noted that the $U_i(j + 1)$ are subtotals of the $U_i(j)$ so that the iteration proceeds very rapidly on an adding machine. The probability that $d \leq k/n$ is $[U_0(n) + U_1(n) + U_2(n) \cdots + U_k(n)]n!n!/(2n)!$.

REFERENCES

[1] W. Feller, "On the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–189.

[2] F. Massey, "A note on the estimation of a distribution function by confidence limits," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 116–119.

[3] F. Massey, "A note on the power of a non-parametric test," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 440–443.

[4] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathématique de l'Université de Moscou*, Vol. 2 (1939), fasc. 2.

[5] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.

A NOTE ON THE SURPRISE INDEX

By R. M. Redheffer

Harvard University

Let $p_m (m = 0, 1, 2, \cdots)$ be a set of probabilities of events E_m, and suppose that the event E_i, with probability p_i, actually occurred. Is the fact that E_i occurred to be regarded as surprising? In a recent article [1] this question is answered by introducing the surprise index S_i,

$$(1) \qquad S_i = (\Sigma p_m^2)/p_i,$$

which gives a comparison between the probability expected and that actually observed.[1] The event is to be regarded as surprising when S_i is large.

The author remarks on the difficulty of computing (1) for the Poisson and binomial distribution. The problem consists in evaluating the numerator, which we shall express here in terms of tabulated functions. The Poisson case leads to Bessel functions, the binomial case to Legendre or hypergeometric functions, and the asymptotic behavior involves square roots only.

1. *The Poisson case.* For the Poisson case we have $p_m = \lambda^m e^{-\lambda}/m!$ so that the generating function is

$$(2) \qquad e^{-\lambda} e^{\lambda x} = \Sigma p_m x^m.$$

Let $x = e^{i\theta}$, then $e^{-i\theta}$; multiply; integrate from 0 to 2π; and simplify slightly to obtain

$$(3) \qquad \Sigma p_m^2 = (e^{-2\lambda}/\pi) \int_0^\pi e^{2\lambda \cos\theta} \, d\theta.$$

---

[1] Cf. also [6].

The integral on the right is recognized[2] as the zero-order Bessel function [2] so that we have

$$(4) \qquad \Sigma p_m^2 = e^{-2\lambda} J_0(-2i\lambda) = e^{-2\lambda} I_0(-2\lambda)$$

as the final answer. The relevant tables are listed on pages 271, 272, and 275 of [5].

2. *The binomial case.* When $p_m = C_m^n p^m q^{n-m}$ with $q = 1 - p$, the value of Σp_m^2 for $p = q = \frac{1}{2}$ is given in the literature [3]; it is the product of the first n odd integers, divided by the product of the first n even integers. For general p,

$$(5) \qquad (q + px)^n = \Sigma p_m x^m$$

is the equation corresponding to (2). Following through the derivation of (3), we get

$$(6) \qquad \Sigma p_m^2 = \frac{1}{2\pi} \int_0^{2\pi} (p^2 + 2pq \cos\theta + q^2)^n \, d\theta$$

which is recognized as the nth order Legendre function [4],

$$(7) \qquad \Sigma p_m^2 = |p - q|^n P_n\left(\left|\frac{p^2 + q^2}{p - q}\right|\right) \qquad (p \ne q).$$

For tables see [5], pages 232–235, 242.

The result (6) is also expressible as a hypergeometric function, and this without intervention of (7). The change of variable $u = p^2 + 2pq \cos\theta + q^2$ leads to

$$(8) \qquad \Sigma p_m^2 = (1/\pi) \int_a^1 u^n (u - a)^{-1/2} (1 - u)^{-1/2} \, du$$

with $a = (p - q)^2$, and letting $u = a + (1 - a)x$ gives an integral which turns out to be [4]

$$\Sigma p_m^2 = (p - q)^{2n} F[-n, \tfrac{1}{2}; 1; -4pq/(p - q)^2].$$

It was brought to the author's attention, by Weaver himself via Mosteller, that (7) is given in Pólya-Szegő, Vol. II, p. 92. There, however, the point of view is to evaluate the integral rather than the sum, and hence the above derivation is the more natural here.

3. *Approximation.* For large values of λ, (4) gives [2]

$$(9) \qquad \Sigma p_m^2 \sim \frac{1}{2\sqrt{\pi\lambda}}.$$

To obtain the asymptotic behavior in the binomial case, note that if the limits of integration in (8) were $0 - 1$, and if the factor $(u - a)^{-1/2}$ were absent, we should have the Beta function $B(n + 1, \frac{1}{2})$. Because u^n emphasizes the region

---

[2] This connection between (3) and the Bessel function was pointed out to the author by by M. V. Cerrillo of M. I. T.

near $u = 1$, this resemblance may be exploited to give (after elementary but tedious calculations)

$$(10) \qquad \pi \Sigma p_m^2 = B(n + 1, \tfrac{1}{2}) + e$$

with

$$0 < e < 2e^{-n^\delta} + (\tfrac{2}{3})[\delta/(1 - a)]^{3/2}$$

whenever $n > a/(1 - a)$. Here δ is any number $< pq$. Picking $\delta = n^{-\theta}$, $\theta < 1$, shows that the error goes to zero almost as fast as $n^{-3/2}$. A similar result may be obtained by the methods of Uspensky.

From (10) we have easily

$$(11) \qquad \Sigma p_m^2 \sim 1/(2\sqrt{\pi n p q}) \qquad\qquad (n \to \infty),$$

which is correct even for $p = q$.

It was pointed out by the referee that (9) and (11) are special cases of the relation

$$\Sigma p_m^2 \sim (\tfrac{1}{2}) \sqrt{\text{variance}}$$

which generally holds whenever the shape of the distribution curve approaches a limit.

REFERENCES

[1] W. WEAVER, "Probability, rarity, interest, and surprise," *The Scientific Monthly*, Vol. 67 (1948), p. 390.

[2] JAHNKE AND EMDE, *Tables of Functions*, 3rd rev. ed., Dover Publications, 1943, p. 149, p. 117.

[3] HALL AND KNIGHT, *Higher Algebra*, Macmillan Co., 1936, p. 148.

[4] WHITAKER AND WATSON, *A Course of Modern Analysis*, 4th ed., Cambridge University Press, 1940, p. 312, p. 293.

[5] FLETCHER, MILLER AND ROSENHEAD, *An Index to Mathematical Tables*, Science Computing Service, Ltd., London, 1946.

[6] T. C. FRY, *Probability and Its Engineering Uses*, D. Van Nostrand Co., 1928, pp. 199–200.

———◆———

APPROXIMATION TO THE POINT BINOMIAL

BY BURTON H. CAMP

Wesleyan University

The following approximation to the sum of the first $(t + 1)$ terms of the point binomial appears to be useful. Let this sum be denoted by S_{t+1}, and let the point binomial be the expansion of $(p + q)^N$; *i.e.*, let

$$(1) \qquad S_{t+1} = p^N + Np^{N-1}q + \cdots + \binom{N}{t} p^{N-t}q^t.$$

Then S_{t+1} is approximately equal to the probability that a unit normal deviate will exceed x, where

$$(2) \qquad x = \frac{\frac{1}{3}\left[\left(\frac{9s-1}{s}\right)\left(\frac{s}{t+1}\frac{q}{b}\right)^{1/3} - \frac{9t+8}{t+1}\right]}{\left[\frac{1}{s}\left(\frac{s}{t+1}\frac{q}{p}\right)^{2/3} + \frac{1}{t+1}\right]^{1/2}}, \qquad s = N - t.$$

This approximation is a corollary to an approximation given by Paulson [1] to the table of the integral of Snedecor's F (Fisher and Yates' $w = e^{2z}$), and the known facts that this integral is an incomplete Beta-function [2] of a simple transform of F, and that S_{t+1} is also an incomplete Beta function of suitable arguments. Paulson's approximation appeared to be quite close. Since it was essentially an approximation to the incomplete Beta function we must now have a similarly close approximation to the point binomial. Therefore two illustrations will suffice.

Example 1. $(.8 + .2)^5$

t	S_{t+1} Approx.	S_{t+1} True	Error
0	.166	.168	−.002
1	.505	.503	.002
2	.801	.797	.004
3	.943	.944	−.001
5	.999	.999	.000

Example 2. $(.9 + .1)^{50}$

t	S_{t+1} Approx.	S_{t+1} True	Error
0	.005	.005	.000
1	.033	.034	−.001
3	.250	.250	.000
5	.617	.616	.001
10	.992	.991	.001

Both these examples involve strongly skewed distributions, one with a small value of N and the other with a fairly large value of N. Considering the amount of computation involved this approximation is much more satisfactory than any other in this author's experience.

REFERENCES

[1] E. PAULSON, "An approximate normalization of the analysis of variance distribution," *Annals of Math. Stat.*, Vol. 13 (1942), p. 233.
[2] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943, p. 115.

A THEOREM ON THE CORRELATION COEFFICIENT FOR SAMPLES OF THREE WHEN THE VARIABLES ARE INDEPENDENT

By C. Chandra Sekar[1]

All India Institute of Hygiene and Public Health, Calcutta

In this note the following theorem will be established:

THEOREM. *If (x_i, y_i) for $i = 1, 2$ and 3 denote three pairs of random values of two independent continuous stochastic variables x and y, r, their correlation coefficient, is given by*

$$(1) \qquad r = \frac{1}{3 s_x s_y} \sum_{i=1}^{3} (x_i - \bar{x})(y_i - \bar{y}),$$

where

$$(2) \qquad \bar{x} = \frac{1}{3} \sum_{i=1}^{3} x_i, \qquad \bar{y} = \frac{1}{3} \sum_{i=1}^{3} y_i,$$

$$s_x^2 = \frac{1}{3} \sum_{i=1}^{3} (x_i - \bar{x})^2, \qquad s_y^2 = \frac{1}{3} \sum_{i=1}^{3} (y_i - \bar{y})^2,$$

and $P(a \leq r \leq b)$ denotes the probability of r taking values in the range $a \leq r \leq b$, then

$$(3) \qquad P\left(-1 \leq r \leq -\frac{1}{2}\right) = P\left(-\frac{1}{2} \leq r \leq \frac{1}{2}\right) = P\left(\frac{1}{2} \leq r \leq 1\right) = \frac{1}{3}.$$

PROOF. If τ_i is defined by

$$(4) \qquad \tau_i = \frac{x_i - \bar{x}}{s_x}, \qquad\qquad i = 1, 2, 3,$$

it is readily seen that the three values of τ are connected by the two relations

$$(5) \qquad \sum_{i=1}^{3} \tau_i = 0, \qquad \sum_{i=1}^{3} \tau_i^2 = 3.$$

Similar conditions exist between the three t_i's defined by

$$(6) \qquad t_i = \frac{y_i - \bar{y}}{s_y}, \qquad\qquad i = 1, 2, 3.$$

The set (τ_1, τ_2, τ_3) can be considered as the Cartesian coordinates of a point in three dimensional space. The conditions (5) restrict the point to a circle. The set (t_1, t_2, t_3) defined by (6) represents a point on the same circle. The correlation coefficient, r, defined in (1) and also given by

$$(7) \qquad r = \frac{1}{3} \sum_{i=1}^{3} \tau_i t_i$$

---

[1] On loan to Population Division, United Nations.

can be regarded as the cosine of the angle θ between the lines joining (τ_1, τ_2, τ_3) and (t_1, t_2, t_3) respectively to the centre of the above-mentioned circle.

The relationships between the τ_i's given by (5) make it necessary for one value of the τ_i's to occur in each of the three non-overlapping intervals $-\sqrt{2}$ to $-\frac{1}{\sqrt{2}}$; $-\frac{1}{\sqrt{2}}$ to $\frac{1}{\sqrt{2}}$ and $\frac{1}{\sqrt{2}}$ to $\sqrt{2}$. Exactly the same conditions hold for the t_i's.[2]

The 6 permutations of τ_1, τ_2, τ_3 in these three intervals correspond to a subdivision of the circle on which the point (τ_1, τ_2, τ_3) lies into 6 equal arcs of 60° each. Every point on any one of these arcs can be shown to correspond, one to one, to the position of τ_i in any one of the intervals; also proceeding along the circle, points on three alternate arcs correspond to the positions of τ_i as it takes on values from the highest to the lowest in this interval and points on the other three correspond to the positions of τ_i as it moves from the lowest to the highest value.

It is clear that when adjacent arcs are combined in pairs dividing the circle into 3 equal arcs of 120°, the probability density function of (τ_1, τ_2, τ_3) is the same on the 3 arcs and is symmetric on each. At any three points on the circle which divide it into three arcs of 120°, the probability density function of (τ_1, τ_2, τ_3) is therefore the same. The same conditions hold for (t_1, t_2, t_3).

It therefore follows that

$$(8) \quad P\left(-\frac{\pi}{3} < \theta \leq \frac{\pi}{3}\right) = P\left(-\frac{2\pi}{3} < \theta \leq -\frac{\pi}{3}, \text{or } \frac{\pi}{3} < \theta \leq \frac{2\pi}{3}\right)$$

$$= P\left(-\pi < \theta \leq -\frac{2\pi}{3} \text{ or } \frac{2\pi}{3} < \theta \leq \pi\right).$$

CORRECTION TO "THE DISTRIBUTION OF EXTREME VALUES IN SAMPLES WHOSE MEMBERS ARE SUBJECT TO A MARKOFF CHAIN CONDITION"

By Benjamin Epstein

Wayne University

In the paper mentioned in the title (*Annals of Math. Stat.*, Vol. 20 (1949), pp. 590–594) I claim to have proved a number of results dealing with the distribution of extreme values in samples of size n drawn at equally spaced intervals from a stationary Markoff process. As Professor W. Feller has kindly pointed

---

[2] This property has been utilised by the author and S. C. Bhoumik to obtain distributions of the correlation coefficient for samples of three, under varying assumptions regarding the distributions of independent variables x and y. The distribution of τ_i or t_i is also of help in working out the distribution of Fisher's g_1 for samples of three. For the distribution of g_1 for samples of three from continuous rectangular distribution, refer to C. Chandra Sekar in *Current Science*, Vol. 13 (1944), pp. 10–11.

out to me in personal correspondence, this is actually not the case. However, the theorems and their proofs remain completely valid in their present form if the observations are drawn from a stochastic process satisfying condition (5) of the paper. This chain condition states that the process be such that

$$\text{Prob } (X_n \leq x \mid X_1 \leq x, X_2 \leq x, \cdots, X_{n-1} \leq x) = \text{Prob } (X_n \leq x \mid X_{n-1} \leq x)$$

is satisfied for all x and for all positive integers n.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Chicago meeting of the Institute, December 27–29, 1950)

1. Cost Functions for Sample Surveys. (Preliminary Report). GARNET E. Mc-CREARY, University of Manitoba and Iowa State College.

Assume: (1) one travels in a rectangular (grid) fashion rather than straight line (airline) path, (2) n random points have a uniform distribution over the region or stratum. Moderate changes in shape of regions have a minor effect on expected distance. Mean airline distance can be prediced from mean grid distance fairly accurately. The following formulas are derived: (1) expected minimum grid distance for $n = 3$ in a square, (2) an upper bound to expected minimum grid distance for all n, (3) expected grid distance for a stratified and unstratified sample, if the path among the points does not reverse a certain direction, (4) expected distance of a random point from (a) the center of the arc of the circle, semicircle or quadrant, (b) any fixed point, inside or outside the rectangular region, (5) mean square distance between any pair of points adjacent in a clockwise direction (6.7 to 9.5 per cent biased upwards over corresponding mean airline distance). Certain conclusions are drawn regarding the most efficient design with respect to total distance. Detailed mileage records of three Iowa farm surveys were compared with theoretical estimates. If the cost is balanced against the losses resulting from errors in estimate, for a particular design, the problem of determining sample size is broached by using Wald's minimax principle.

2. On a Preliminary Test for Pooling Mean Squares in the Analysis of Variance. A. E. PAULL, Abitibi Power and Paper Company, Limited, Toronto, Canada.

The consequences of performing a preliminary F-test in the analysis of variance is described. The use of the 5% or 25% significance level for the preliminary test results in disturbances that are frequently large enough to lead to incorrect inferences in the final test. A more stable procedure is recommended for performing the preliminary test, in which the two mean squares are pooled only if their ratio is less than twice the 50% point.

3. Estimation for Sub-Sampling Designs Employing the County as a Primary Sampling Unit. EMIL H. JEBE, Iowa State College and North Carolina State College.

This paper summarizes a study of the application of various two-stage designs including the estimation procedures for providing state estimates of agricultural items in North Carolina. Among the principal objectives of the investigation were (1) the comparison of the efficiency of selection of the primary units with equal and with unequal probabilities, and (2) assessment of the relative contributions of the between primary sampling unit and within primary sampling unit error components to the total sampling error. Examination of several linear and ratio estimates indicates a number of advantages for a particular ratio estimate.

4. The Probability Distribution of the Number of Isolates in a Social Group.
Leo Katz, Michigan State College.

Each of the N members of a well-defined social group is asked to name d others with whom he would prefer to be associated in some specified activity. Under the null hypothesis, his choices are randomly distributed. An isolate is an individual who is not chosen by any of the other members of the group. The probability of exactly i isolates in the group is then given by

$$P_i = \Sigma_{j=1}^{N-1-d} (-1)^{i+j} C(j, i) C(N, i) [C(N - i, d)]^j [C(N - 1 - i, d)]^{N-j} [C(N - 1, d)]^{-N},$$

where $C(N, n) = {}_N C_n$, the binomial coefficient. This expression for P_i is somewhat unwieldy. It is further shown that this probability function is asymptotically a binomial p.f., $P_i' = C(n, i) p^i (1 - p)^{n-i}$, where

$$p = N[(N - 1 - d)/(N - 1)]^{N-1} - (N - 1)[(N - 1 - d)/(N - 1)][(N - 2 - d)/(N - 2)]^{N-2}$$

and $np = N[(N - 1 - d)/(N - 1)]^{N-1}$. The approximation is very good even for moderately small values of N.

5. Estimating Population Size Using Sequential Sampling Tagging Methods.
Leo A. Goodman, University of Chicago.

Let $[n_i]$ be a sequence of positive integers and let $S(L, n_i)$ denote the procedure whereby (1) n_1 elements are drawn at random from a population P, then tagged to distinguish them from the remaining elements, and replaced in P, (2) n_2 elements are drawn from P, the number of tagged elements appearing is observed, the n_3 elements are then tagged and replaced in P, (3) $\cdots$, this process is halted when at least $L > 0$ tagged elements have appeared. Given $S(L, n_i)$, there exists a minimum variance unbiased estimator (m.v.u.e.) of the number N of elements in P which may be determined as the quotient of two determinants and simplified, by combinatorial methods, in special cases. If $[n_i]$ is bounded, as N approaches infinity, the limiting distribution of t^2/N, where t is the total number of elements drawn before the procedure ceases, is χ^2 with $2L$ degrees of freedom. Hence the asymptotic m.v.u.e. of N, confidence intervals and tests of hypotheses for N may be obtained as well as the approximate fiducial distribution of N. Similar results may be obtained for the more general cases where (a) information concerning size of some subclasses in P is used and (b) where taggings may or may not be differentiated. The $S(L, n_i)$ compares favorably with other procedures considered.

6. Application of the Distribution of a Linear Form in Chi-square Variates.
Arthur Grad and Herbert Solomon, Office of Naval Research, Washington, D. C.

The probability of hitting a target depends both on the accuracy with which the position of the target is known and the dispersion of the weapon about the point of aim. Under the assumption that each of these errors has a bivariate Gaussian distribution with known covariance matrix, $\| \sigma(p) \|$ for position prediction error and $\| \sigma(a) \|$ for aiming error, about the point of aim (predicted position), the probability, P, of hitting the target with a weapon having a radius of effectiveness R is shown to be $P = Pr\{k_1^2 x_1^2 + k_2^2 x_2^2 \leq R^2/C^2\}$, where $k_1^2 = [\sigma_{11}(p) + \sigma_{11}(a)]/C^2$, $k_2^2 = [\sigma_{22}(p) + \sigma_{22}(a)]/C^2$, $C^2 = \sigma_{11}(p) + \sigma_{22}(p) + \sigma_{11}(a) + \sigma_{22}(a)$, and x_i^2 is a chi-square variate with 1 degree of freedom. When $\sigma_{12}(p) = \sigma_{12}(a) = 0$, then the chi-square variates are independent. If not, a linear transformation exists such that $z = k_1^2 x_1^2 + k_2^2 x_2^2 = l_1^2 y_1^2 + l_2^2 y_2^2$, where $l_1^2 + l_2^2 = k_1^2 + k_2^2$ and y_1 and y_2 are independently dis-

tributed chi-square variates each having one degree of freedom. It is then demonstrated that $P = 2k_1k_2 \int_0^t e^{-z}I_0[z(1 - 4k_1^2k_2^2)^{\frac{1}{2}}]\,dz$, where $t = R^2/4C^2k_1^2k_2^2$, when the chi-square variates are independent; in case of dependence, k_i should be replaced by l_i. A table was constructed which covers the entire range of the parameters.

7. A Large Sample t-statistic which Is Insensitive to Nonrandomness. JOHN E. WALSH. The Rand Corporation.

Most of the well known significance tests and confidence intervals for the population mean are based on the assumption of a random sample. This paper considers how the significance levels and confidence coefficients of the commonly used class of tests and intervals based on the standard Student t-statistic are changed when the random sample requirement is violated and the number of observations is large. It is found that even a slight deviation from the random sample situation can result in a substantial significance level and confidence coefficient change. Thus this class of tests and confidence intervals would seem to be of questionable practical value for large sets of observations. Large sample tests and confidence intervals for the mean which are not sensitive to the random sample requirement are obtained for a situation of practical interest by development of a special type of t-statistic. These results are as efficient (asymptotically) as those based on the standard t-statistic for the case of a random sample.

8. Conditional Expectation and Convex Functions. E. W. BARANKIN, University of California, Berkeley.

The inequality $E\psi(E(f \mid \cdot)) \leq E\psi(f)$, (where the conditional expectation is taken with respect to a function t) with f a real- (or complex-) valued function on the fundamental space, was shown by Blackwell to hold in the case $\psi(z) = \mid z \mid^2$, and by the present author to hold in the case $\psi(z) = \mid z \mid^s$, $s \geq 1$ (*Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110, and Vol. 21 (1950), pp. 280–284, respectively). More recently Hodges and Lehmann (*Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197) proved the inequality in the case of f a function to $\mathcal{E}^k$ (Euclidean k-space) and ψ a finite, convex, real-valued function on $\mathcal{E}^k$. Now, both Blackwell and this author exhibited the above inequality, in their cases, as (obvious) consequences of the more fundamental relation: $\psi(E(f \mid \tau)) \leq E(\psi(f) \mid \tau)$ *for almost all points* τ *in the range of* t. The work of Hodges and Lehmann, however, leaves open the question whether or not the latter inequality holds in the more general case. In the present note this almost-everywhere inequality is established for f to $\mathcal{E}^k$ and ψ convex. The first inequality then obtains by integration.

9. Transformation Parameters. MELVIN P. PEISAKOFF, The Rand Corporation.

Location, scale, and location-scale parameters are examples of *transformation parameters*. Transformation parameters are defined by: (1) the parameter space is a group, (2) the sample space can be factored into the same group and an arbitrary space, (3) the random variable associated with each parameter point, θ, can be generated by drawing from the population associated with the unit of the parameter space and left multiplying the group component of the sample by θ. *Decision function theory* is investigated when the decision space and the cost function are of a special intuitively appealing form. The formulation is broad enough to include sequential analysis. Minimax decision functions are found. Also investigated is *testing and confidence region theory*, using extensively the results on decision functions. Both simple and composite hypotheses are treated. Finally, (Fisher) *information theory* is examined. It is shown that modifications are necessary if information theory is to be useful in estimation problems. One modification is suggested. This modification en-

larges the class of standard estimators to include each estimator which is minimax with respect to a certain risk function determined by the estimator itself. The approach is generalized to include inequalities for the mean square error other than the information inequality.

10. A Generalization of the Neyman-Pearson Fundamental Lemma. HENRY SCHEFFÉ, Columbia University.

Given $m + n$ real integrable functions $f_1(x), \cdots, f_m(x), h_1(x), \cdots, h_n(x)$ of a point x in a Euclidean space R, a real function $\varphi(y_1, \cdots, y_n)$ of n real variables, and m constants $c_1, \cdots, c_m$, the problem is to consider the existence of, and to find necessary conditions and sufficient conditions on, a set S maximizing $\varphi\left(\int_S h_1\, dx, \cdots, \int_S h_n\, dx\right)$ subject to the m side conditions $\int_S f_i\, dx = c_i$. In some applications the values of the vector

$$\left(\int_S h_1\, dx, \cdots, \int_S h_n\, dx\right)$$

may also be restricted to a given set. A statistical example in which $\varphi(y_1, \cdots, y_n) = \prod_{i=1}^{n} y_i$ arose in an unpublished paper of Isaacson. The methods of the present paper are suggested by those of an unpublished paper of Dantzig and Wald. Under certain regularity conditions the inequalities appearing in the Neyman-Pearson lemma are replaced by $\sum_{i=1}^{n} a_i^S h_i(x) - \sum_{j=1}^{m} k_j f_j(x) \geq 0$ (a.e. in S), ≤ 0 (a.e. in $R - S$). Here a_i^S and k_j are constants with $a_i^S = \partial \varphi / \partial y_i$ evaluated at $(y_1, \cdots, y_n) = \left(\int_S h_1\, dx, \cdots, \int_S h_n\, dx\right)$.

11. Nonparametric Estimation V, Sequentially Determined Statistically Equivalent Blocks. D. A. S. FRASER, University of Toronto.

In 1943 Wald gave a method for constructing tolerance regions for continuous multivariate distributions. Tukey generalized Wald's procedure and then interpreted the results for discontinuous distributions. In this paper a further generalization of the method is given by which statistically equivalent blocks can be determined sequentially; that is, the particular function used to cut off a block may depend on the shape or structure of previously selected blocks. The results are also extended to the case of discontinuous distributions. Possible advantages for the practitioner are discussed.

12. A Bayes Approach to a Quality Control Model. M. A. GIRSHICK AND HERMAN RUBIN, Stanford University.

A machine producing items of quality characteristic x can be in one of four states. In state $i = 1, 2$ the machine is in production and is characterized by a density $f_i(x)$. In state $j = 3, 4$ the machine is in repair having come from state $j = 2$. When the machine is in state 1 there is a probability g that in the next time unit it enters state 2, remaining in state 2 until brought to repair by some rule R based on observations. The income from items of quality x is $V(x)$; repair cost per unit time in state $j = 3, 4$ is c_j. A rule R^* is Bayes if it maximizes $\lim I_N$ as $N \to \infty$ where I_N is the expected income per unit time in N time units. It is proved that for 100% inspection, R^* states that sampling is to continue as long as $Z_n < a$ and sampling is to terminate and the machine placed in repair when

$Z_n \geq a$, where $Z_n = y_n(1 + Z_{n-1})$, $Z_0 = 0$ and $y_n = f_2(x_n)/(1 - g)f_1(x_n)$. R^* is also obtained in case inspection costs are taken into account. It is shown that the above Markoff process approaches a stable distribution and the required integral equations are derived.

13. On the Translation Parameter Problem for Discrete Variables. DAVID BLACKWELL, Stanford University.

Let $x = (x_1, \cdots, x_N)$ be a vector chance variable, let $y = x + h\epsilon$, where $\epsilon = (1, \cdots, 1)$ and h is an unknown constant, and let $t = t(y)$ be any function of y, considered as an estimate for h when y is observed. Let $f(d)$ be any function of a real variable d, considered as the loss to the statistician when the error of estimate is d, so that the risk from an estimate t is $R_t(h) = Ef[h - t(x + \epsilon h)]$. Extending the work of Pitman, Girshick and Savage have exhibited an estimate t^* for which $R_t^*(h) = R$ independent of h, and have shown that t^* is minimax. It is shown here that if x assumes only a finite number of values $v_i = (n_{i1}, \cdots, n_{iN})$ and each n_{ij} is an integer, and if $f(d)$ is strictly convex and assumes its minimum value, then t^* is admissible and is in fact the unique minimax estimate. Two examples in which t^* is not admissible are given. A closely related fact is that if S is a closed bounded strictly convex subset of n-space intersecting the line $x_1 = \cdots = x_n$ at the single point $(w, \cdots, w)$, then the only sequence $\{z_m\}$, $-\infty < m < \infty$, for which $P_m = (z_{m+1}, \cdots, z_{m+n}) \in S$ for all m is $z_m = w$ for all m.

14. On Ratios of Certain Algebraic Forms. ROBERT V. HOGG, State University of Iowa.

Let x and y be random variables having a continuous cumulative distribution function, and let $M(u, t) = E[\exp(ux + ty)]$ exist in the neighborhood of the origin of the u, t plane. Subject to certain conditions a necessary and sufficient condition for the stochastic independence of y and x/y is $(\partial^k/\partial u^k)M(0, t) \equiv K_k(\partial^k/\partial t^k)M(0, t)(k = 0, 1, 2, \cdots)$, where K_k is evaluated by setting $t = 0$. This result is used in the study of certain ratios of quadratic and linear forms. In dealing with the quadratic forms, the sample arises from a normal population with mean zero. A necessary and sufficient condition is determined for the stochastic independence of Q_2 and Q_1/Q_2, where essentially $Q_1 = a_1 x_1^2 + \cdots + a_n x_n^2$ and $Q_2 = b_1 x_1^2 + \cdots + b_n x_n^2$. In the linear case however, the distribution is unspecified. Then it is found that the requirement of the stochastic independence of L_2 and L_1/L_2 implies that the sample arose from a gamma type distribution. Here $L_1 = a_1 x_1 + \cdots + a_n x_n$ and $L_2 = x_1 + \cdots + x_n$.

15. The Economics of Sampling. NORMAN RUDY, Sacramento State College.

An optimum single sampling plan for acceptance inspection of attributes is developed by the method of minimizing the maximum risk. The first application is to warehouse or surveillance inspection, in which the value of a good item, g, and the cost of a bad item, b, define a breakeven quality, p_0. It is shown that under these conditions, and with sampling cost a linear function of sample size, s, tn, the optimum sample size is approximately equal to $[(.085 \text{ lot size})/t]^{2/3} (bg)^{1/3}$, the optimum acceptance number is approximately equal to np_0, and the $\min_{n,a} \max_p$ of the risk is approximately equal to $s + .58(tbg)^{1/3} (\text{lot size})^{2/3}$. The more general case, where the breakeven quality p_0 is determined by trade practice or contract, is also worked out, but cannot be presented in completely analytic form. A simple table involving the quotient of the normal integral and the normal density is required. Given this and the cost parameters of the situation, then the sample size and the acceptance number which minimize the maximum risk are determined from relatively simple expressions.

16. Exact Tests of Serial Correlation Using Noncircular Statistics. G. S. WATSON, University of Cambridge, AND J. DURBIN, London School of Economics.

The paper shows how noncircular statistics for testing hypotheses of serial independence may be constructed for which exact distributions can be obtained using results given by R. L. Anderson ("Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13). The statistics are derived by throwing away a small amount of relevant information. As an example the statistic

$$c_1 = (x_1 x_2 + \cdots + x_{m-1} x_m + x_{m+1} x_{m+2} + \cdots + x_{2m-1} x_{2m}) \Big/ \sum_1^{2m} x_i^2$$

may be used for testing independence in a series of $2m$ observations whose mean is known to be zero. The quadratic form in the numerator of c_1 is based on a matrix whose roots are pair-wise equal, so that the distribution of c_1 when the x's are normal with the same variance is known from the results of R. L. Anderson. Tests of the errors in certain regression models may be made by fitting separate regressions to the two halves of the series and substituting the residuals in expressions similar to c_1. Exact tests can be obtained in this way for polynomial regressions, one-way, two-way etc. classifications, and periodic regressions. The statistics appear to have power comparable with that of the related circular statistics against alternative hypotheses specified by a stationary Markoff process. In many cases occurring in practice, however, serial correlation of the errors will be due to systematic behaviour arising from the inadequacy of the theoretical model to represent the true relationship. The statistics proposed will often be preferable to circular statistics in such cases.

17. Stochastic Difference Equations with a Continuous Time Parameter. (Preliminary Report). S. G. GHURYE, University of North Carolina.

Given a discrete sequence of observations ordered equidistantly in time, it is often assumed that this discrete process is explained by a stochastic difference equation with a purely random "disturbance". However, this observed discrete process might be the result of observations on a stochastic process $X(t)$ in which t is not discrete, but continuous. Is it possible to have a process $X(t)$, defined for t real, such that given any real t_0 and any real $h > 0$, the sequence $\{X(t_0 \pm jh)\}$, $j = 0, 1, \cdots$, satisfies the equation

$$X(t_0 + [j + p]h) + \alpha_1(h)X(t_0 + [j + p - 1]h) + \cdots + \alpha_p(h)X(t_0 + jh) = \delta(t_0 + jh),$$

δ being a linear function of mutually independent random variables having a common c.d.f. which is independent of h? The cases $p = 1$ and $p = 2$ are dealt with in detail, and the possible forms of such processes derived; the further problem for any p, as also for a system of equations, is being considered. It is also proposed to tackle the problems of estimation and testing which arise in this connection.

18. Nonsequential Problems in the Case of k Hypotheses. (Preliminary Report). HERMAN CHERNOFF, University of Illinois.

Suppose that there are k possible simple hypotheses $H_1, H_2, \cdots, H_k$ and a possibly infinite set of actions may be taken. To a decision function there corresponds a vector $\rho = (\rho_1, \rho_2, \cdots, \rho_k)$ where ρ_i is the risk if H_i is true. The closure of the range of ρ is convex in the nonatomic case and in the randomized case. In the randomized case the closure of the range of ρ is the convex hull of the closure of the range of ρ in the nonrandomized case. (The randomized case is that one where a number is selected at random from the unit interval before an action is taken.) The range of ρ is closed under suitable closure conditions on the range of the weight function.

19. The Moments of a Multinormal Distribution after One-sided Truncation of Some or All Coordinates. Z. W. Birnbaum and Paul L. Meyer, University of Washington.

Let $X = (X_1, X_2, \cdots, X_p)$ be a multinormal random variable with given first and second moments and the probability density $f(X_1, X_2, \cdots, X_p)$. The random variable $Y = (Y_1, Y_2, \cdots, Y_p)$ is said to be obtained from X by truncation to the set $X_i \geq \tau_i$, $i = 1, 2, \cdots, p$, if its probability density is $g(Y_1, Y_2, \cdots, Y_p) = Cf(Y_1, Y_2, \cdots, Y_p)$ for $Y_1 \geq \tau_1, Y_2 \geq \tau_2, \cdots, Y_p \geq \tau_p$, and $g(Y_1, Y_2, \cdots, Y_p) = 0$ elsewhere. The problem considered is to determine the mathematical expectations $E(Y_i^m Y_j^n)$. Explicit formulae are obtained for the first and second moments $E(Y_i)$ and $E(Y_i Y_j)$, and recursion formulae are given for the general case. (Research done under the sponsorship of the Office of Naval Research.)

20. An Algorithm for the Determination of all Solutions of a Two-Person Zero Sum Game with a Finite Number of Strategies. H. Raiffa, G. L. Thompson, and R. M. Thrall, University of Michigan.

Consider a zero-sum two-person game in which each player has a finite number of strategies. A computational procedure is given for finding the value of the game and all optimal basic strategies for each player. The basic computations required are evaluation of linear forms and solution of linear equations in one unknown. This method, based on geometric reasoning, is a step by step process with no more stages than the total number of strategies for the two players.

21. A Note on the Convolution of Uniform Distributions. Edwin G. Olds, Carnegie Institute of Technology.

Let X_i be independent random variables with probability density functions $[\epsilon(X_i) - \epsilon(X_i - a_i)]/a_i$, where $\epsilon(x - c)$ is unity for $x \geq c$ and zero elsewhere. This paper gives a simple proof that the probability density function for $S = \Sigma_1^n x_i$ is

$$[S^{n-1}\epsilon(S) - \Sigma_1^n (S - a_i)^{n-1}\epsilon(S - a_i) + \Sigma_{i<j} (S - a_i - a_j)^{n-1}\epsilon(S - a_i - a_j)$$

$$- \cdots + (-1)^n (S - \Sigma a_i)^{n-1}\epsilon(S - \Sigma a_i)]/(n - 1)!\Pi_1^n a_i.$$

A sufficient condition for the asymptotic normality of S is $0 < \alpha \leq a_i \leq \beta$ (finite). For the special case where $a_{i+1} = ra_i$ the necessary and sufficient condition for asymptotic normality is $r = 1$. For $0 \leq r \leq 0.5$ or $r \geq 2$ the probability that S will be outside the interval $\mu_S \pm 3\sigma_S$ is zero. From the Edgeworth Series for the distribution function for the standardized sum it follows that $F(-3) \doteq 0.00135 - 0.004[\Sigma a_i^4/(\Sigma a_i^2)^2]$ where the bracketed expression takes its minimum value n^{-1} when all of the a_i's are equal. These results are useful in connection with the problem of random assembly.

22. On the Consistency of Certain Estimates of the Linear Structural Relation. Elizabeth L. Scott, University of California, Berkeley.

Let $\{x_i, y_i\}$ denote n independent pairs of observations on x, y where $x = \xi + u$ and $y = \alpha + \beta\xi + v$ with ξ, u and v random variables with finite variances, $E(u) = E(v) = 0$ and ξ independent of the pair u, v. *Procedure* (1): Fix $a \leq b$ such that

$$P\{x \leq a\} > 0, \qquad P\{x > b\} > 0.$$

Let X_1, Y_1 stand for the arithmetic mean of the x_i's and y_i's, respectively, for $x_i \leq a$ and X_2, Y_2 for those for which $x_i > b$. As an estimate of β, consider, say, $b_1 =$

$(Y_2 - Y_1)/(X_2 - X_1)$. *Procedure* (2): Let X_1, Y_1 denote arithmetic mean of x_i's and y_i's, respectively, for which x_i is one of the r smallest of the x_i's and X_2, Y_2 for those for which x_i is one of the s largest, with r, s preassigned, $r < n - s + 1$. The corresponding estimate of β is, say, b_2 defined as above. Let (μ, ν) denote the shortest interval such that $P\{\mu \leq u \leq \nu\} = 1$. THEOREM 1. *In order that b_1 preserve the property of being a consistent estimate of β irrespective of the value of β, $-\infty < \beta < \infty$, it is n.a.s. that $P\{a - \nu < \xi \leq a - \mu\} = P\{b - \nu < \xi \leq b - \mu\} = 0$.* Now let $r = p_1 n$, $s = p_2 n$ and m, M be the corresponding percentile points such that $P\{\xi \leq m\} = p_1$ and $P\{\xi > M\} = p_2$. THEOREM 2. *If $n \to \infty$ while p_1 and p_2 are held constant, the n.a.s. condition that b_2 preserve the property of being a consistent estimate of β irrespective of the value of β, $-\infty < \beta < \infty$, is that $P\{m - \nu < \xi \leq m - \mu\} = P\{M - \nu < \xi \leq M - \mu\} = 0$.* Similar estimates were considered, for $p_1 = p_2 = \frac{1}{2}$, u and v independent, by A. Wald (*Annals of Math. Stat.*, Vol. 11 (1940), pp. 295–297) who showed sufficiency.

23. A 3-decision Problem Concerning the Mean of a Normal Population. R. R. BAHADUR, University of Chicago.

Given n independent observations x_1, x_2, $\cdots$, x_n from a normal population having an unknown mean $\theta\sigma$ and unknown variance σ^2, suppose that the statistician is asked to say whether the unknown mean is $>c$ or $\leq c$ where c is a given constant (which is supposed henceforth to be zero), or to say that he would rather reserve judgement on the matter. In the present problem (which was suggested by Professor R. C. Bose as a modification of the problems considered in "The Problem of the Greater Mean," [R. R. BAHADUR AND H. ROBBINS, *Annals of Math. Stat.*, Vol. 21 (1950), pp. 469–487]), reserving judgement is considered to be undesirable, and the possibility of doing so is admitted only for the purpose of reducing the probability of the statistician making an incorrect assertion. For any procedure d which associates each sample with one of the three decisions "assert $\theta > 0$", "assert $\theta \leq 0$", and "reserve judgement", let $a(d \mid \theta\sigma, \sigma) = Pr.$("incorrect assertion" using $d \mid \theta\sigma, \sigma$), $b(d \mid \theta\sigma, \sigma) = Pr.$("reserve judgement" using $d \mid \theta\sigma, \sigma$), and set $\alpha(d \mid \theta) = \sup_\sigma \{[a(d \mid \theta\sigma, \sigma) + a(d \mid -\theta\sigma, \sigma)]/2\}$,

$$\beta(d \mid \theta) = \sup_\sigma \{[b(d \mid \theta\sigma, \sigma) + b(d \mid -\theta\sigma, \sigma)]/2\}.$$

The class of procedures $\{d_\tau^*\}$ is defined as follows: for any τ, $0 \leq \tau \leq \infty$, $d_\tau^* =$ "assert $\theta > 0$ if $\bar{x} > \tau s$, assert $\theta \leq 0$ if $\bar{x} \leq -\tau s$, and reserve judgement otherwise", where $\bar{x} = n^{-1}\Sigma_1^n x_i$ and $s^2 = n^{-1}\Sigma_1^n(x_i - \bar{x})^2$. One of the results obtained concerning the class $\{d_\tau^*\}$ is as follows. *Corresponding to any d there exists a d_τ^* such that $\alpha(d_\tau^* \mid \theta) \leq \alpha(d \mid \theta)$ and $\beta(d_\tau^* \mid \theta) \leq \beta(d \mid \theta)$ for all θ.* In particular, given p, $(0 < p < \frac{1}{2})$, there (evidently) exists a $\tau(p)$, $(0 < \tau(p) < \infty)$, such that $\sup_\theta \{\alpha(d_{\tau(p)}^* \mid \theta)\} = p$, and if d is any other procedure such that $\sup_\theta \{\alpha(d \mid \theta)\} \leq p$, then $\beta(d \mid \theta) \geq \beta(d_{\tau(p)}^* \mid \theta)$ for all θ. These results provide a justification of the manner in which the two-sided t test of a normal mean is sometimes used in practice.

24. Consistent Estimate of the Slope of a Linear Structural Relation. J. NEYMAN, University of California, Berkeley, AND CHARLES M. STEIN, University of Chicago.

Let Z_n denote the system of $8n$ independent pairs of measurements (X_{ik}, Y_{ik}), for $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, 8$, of two nonobservable random variables ξ_{ik} and $\eta_{ik} = \alpha \csc\beta - \xi_{ik} \cot\beta$, where α and β are constants. Variable ξ_{ik} is nonnormal. It is assumed that any nonnormal components of the errors of measurement $X_{ik} - \xi_{ik}$ and $Y_{ik} - \eta_{ik}$ are mutually independent, independent of ξ_{ik} and of the normal components of the errors. The normal components of errors may be correlated but as a pair are independent of ξ_{ik}. For every $n \geq 4$, let $m(n)$ be the greatest integer not exceeding $\sqrt{n}$. Let $\Delta(n) = \pi/(m(n) - 1)$ and $b_{n,j} = -\pi/2 + (j - 1)\Delta(n)$, for $j = 1, 2, \cdots, m(n)$. For every b, $|b| \leq \pi/2$ and for

$i = 1, 2, \cdots, n$ let $A_i = \exp\{-\frac{1}{2}[(X_{i1} - X_{i2} + X_{i3} - X_{i4})\cos b + (Y_{i1} - Y_{i2} + Y - Y_{i4})\sin b]^2 - \frac{1}{2}(X_{i1} + X_{i3} + X_{i4})^2]\}$, $B_i = \exp\{-\frac{1}{2}(Y_{i1} - Y_{i2} + Y_{i7} - Y_{i8})^2\}$ $C_i = \exp\{-\frac{1}{2}(Y_{i1} - Y_{i4} + Y_{i6} - Y_{i7})^2\}$, $D_i = \exp\{-\frac{1}{2}(Y_{i3} - Y_{i4} + Y_{i6} - Y_{i8})^2\}$, and finally, $G(b, Z_n) = [\Sigma_{i=1}^n A_i(B_i - 2C_i + D_i)]/n$. Let $g(Z_n)$ be the smallest of the $m(n)$ values of the function $G(b, Z_n)$ computed for $b_{n1}, b_{n2}, \cdots, b_{nm(n)}$ and let $T(Z_n)$ denote the smallest of the b_{nj} for which $G(b_{nj}, Z_n) = g(Z_n)$. THEOREM. *As* $n \to \infty$, *the function* $T(Z_n)$ *thus defined is a consistent estimate of* β. The present problem grew out of the problem of identifiability of β studied by Olav Reiersøl (*Econometrica*, Vol. 18 (1950), pp. 375–389). The results obtained here represent a generalization of the previous results of one of the authors presented at the I.M.S. meeting in Boulder, Colorado, as the Second Rietz Memorial Lecture, September, 1949.

25. A Remark on Almost Sure Convergence. MICHEL LOÈVE, University of California, Berkeley.

A criterion for almost sure convergence is given. It contains criteria of Kolmogorov, Marcinkiewicz, and P. Lévy.

26. A Significance Test for Differences Among Ranked Treatments in an Analysis of Variance. D. B. DUNCAN, Virginia Polytechnic Institute.

Given a set of n treatment means (or totals) $x_1, x_2, \cdots, x_n$, it is often desired to decide whether each of the differences $x_j - x_i$ is significant, that is, whether each of the hypotheses $H: \mu_j > \mu_i$, $i, j = 1, 2, \cdots, n$, $i \neq j$ can be accepted. A test is obtained for this purpose under the conditions which usually apply or are taken to apply in many analyses of variance, namely that $x_1, x_2, \cdots, x_n$ is a random sample from n normal populations with means $\mu_1, \mu_2, \cdots, \mu_n$, respectively, and a common unknown variance σ^2 for which the common form of independent estimate s^2 based on n_2 degrees of freedom is available. In approaching the problem the complete Wald multiple decision function form of analysis is found to be too unwieldy for a general case and is waived in favor of a simpler set of requirements. These state that an α level test should provide likelihood ratio tests as closely as possible for each of the $_nC_r$ hypotheses that any combination of r of the treatment means are equal. Also satisfactory upper limits should be placed on the significance level of the whole test with respect to each of these particular $_nC_r$ hypotheses. The test obtained satisfies the given requirements better than other currently available procedures. It consists of a fairly simple sequence of range-like tests followed by variance tests which are presented in detail together with examples.

27. On Information and Sufficiency. S. KULLBACK, George Washington University, AND R. A. LEIBLER, Washington, D. C.

For probability spaces $(\mathbf{X}, \mathbf{S}, \mu_i)$, $i = 1, 2$, and probability measures λ, μ_1, μ_2 absolutely continuous with respect to each other in pairs, f_i, $i = 1, 2$, is defined by

$$\mu_i(E) = \int_E f_i(x) \, d\lambda(x) \quad \text{for all} \quad E \in \mathbf{S}.$$

Then $I_{1:2}(E) = [1/\mu_1(E)] \int_E f_1(x)[\log f_1(x) - \log f_2(x)] \, d\lambda(x)$ for $\mu_1(E) > 0$, and $I_{1:2}(E) = 0$ for $\mu_1(E) = 0$, is defined as the mean information for discrimination between H_1 and H_2 per observation from $E \in \mathbf{S}$ for μ_1, where H_i is the hypothesis that x is selected from the population with probability measure μ_i. $J_{12}(E)$, the divergence between the populations in E, is defined as $I_{1:2}(E) + I_{2:1}(E)$ or

$$J_{12}(E) = \int_E [f_1(x)/\mu_1(E) - f_2(x)\mu_2(E)][\log f_1(x) - \log f_2(x)] \, d\lambda(x).$$

Properties of I and J are considered and the relations of I to the information notions of Fisher, Shannon and Wiener and J to Mahalanobis' generalized distance are noted. In particular it is proved that a transformation T never increases $I_{1:2}(\mathbf{X})$ and a necessary and sufficient condition that T leave $I_{1:2}(\mathbf{X})$ unchanged is that T be a sufficient statistic.

28. **Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes.** T. W. Anderson, Columbia University, and D. A. Darling, University of Michigan.

The statistical problem treated is that of testing the hypothesis that a sample of n independent, identically distributed random variables have the common continuous distribution function $F(x)$. If $F_n(x)$ is the empirical cumulative distribution function and $\psi(x)$ is some nonnegative weight function $(0 \leq x \leq 1)$, we consider

$$K_n = n^{\frac{1}{2}} \sup_{-\infty < x < \infty} \{| F(x) - F_n(x) | \psi^{\frac{1}{2}}[F(x)]\}$$

and $W_n^2 = n \int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 \psi[F(x)] \, dF(x)$. For suitable choices of ψ these tests have been considered by Kolmogorov, Cramér, von Mises, Smirnov, and others. A unified method for calculating the limiting distributions of K_n and W_n^2 is developed by reducing them to corresponding problems in stochastic processes, which in turn lead to more or less classical eigen-value and boundary value problems for special classes of differential equations. For certain weight functions we give explicit limiting distributions. For $\psi = 1$ we obtain, e.g., the Kolmogorov distribution and the ω^2 distribution of Smirnov and von Mises for K_n and W_n^2, respectively. By courtesy of the numerical analysis section of the Rand Corporation a tabulation of the ω^2 distribution has been prepared. (This work was supported by the Rand Corporation.)

29. **The Effect of Preliminary Tests of Significance on the Size and Power of Certain Tests of Univariate Linear Hypotheses with Special Reference to the Analysis of Variance. (Preliminary Report).** Robert E. Bechhofer, Columbia University.

Let $X_1, \cdots, X_q; Y_1, \cdots, Y_r; Z_1, \cdots, Z_s$ be normally and independently distributed with means $0, \cdots, 0; \mu_1, \cdots, \mu_r; \nu_1, \cdots, \nu_s$, respectively, and variance σ^2. The null hypothesis is $H_0: \nu_1 = \cdots = \nu_s = 0$. The standard test (T_1) of H_0 is an F-test involving $\Sigma_{i=1}^{s} Z_k^2 / \Sigma_{i=1}^{q} X_i^2$. If $\mu_1 = \cdots = \mu_r = 0$, a more powerful test (T_2) of H_0 is an F-test involving $\Sigma_{k=1}^{s} Z_k^2 / (\Sigma_{i=1}^{q} X_i^2 + \Sigma_{j=1}^{r} Y_j^2)$. However, if $\Sigma_{j=1}^{r} \mu_j^2 / \sigma^2$ should be large, T_2 would have low power. When the statistician believes (based on past experience) that the μ's equal zero, but wishes to protect himself against the possibility that they do not, he can use a preliminary F-test (T_0), i.e., he pools (uses T_2) or does not pool (uses T_1) accordingly as $\Sigma_{j=1}^{r} Y_j^2 / \Sigma_{i=1}^{q} X_i^2$ is less than or greater than some preassigned constant. The power of the composite test $[T = (T_0 \text{ plus } T_1 \text{ or } T_2)]$ depends on q, r, s; the levels of significance $\alpha_0, \alpha_1, \alpha_2$ associated with T_0, T_1, T_2, respectively; and $\lambda_2 = \Sigma_{j=1}^{r} \mu_j^2 / 2\sigma^2$ (the nuisance parameter) and $\lambda_3 = \Sigma_{k=1}^{s} \nu_k^2 / 2\sigma^2$. Formulae are derived for the size (Type I error) and power of T. The behavior of the size and power as a function of λ_2 and λ_3 is characterized. It is shown that certain choices of $\alpha_0, \alpha_1, \alpha_2$ yield tests T which have desirable properties. (Part of this work was carried out under the sponsorship of the Office of Naval Research.)

30. **The Exact Distribution of the Extremal Quotient.** E. J. Gumbel, New York, and L. H. Herbach, Columbia University.

The distribution of the extremal quotient q (the ratio of the largest value x_n to the smallest x_1 of n independent observations taken from the same distribution), is obtained in four stages, three special cases: (1) $x_1 \geq 0, x_n \geq 0, q \geq 1$. (2) $x_1 \leq 0, x_n \leq 0,$

$0 \leq q \leq 1$. (3) $x_1 \leq 0$, $x_n \geq 0$, $q \leq 0$, culminating in the general case: (4) $-\omega_1 \leq x_1 \leq x_n \leq \omega_2$, $-\omega_2/\omega_1 \leq q < \infty$. The common procedure in the first three cases is to integrate out the extreme from the joint distribution of one extreme and the extremal quotient. Geometric considerations give the appropriate regions of integration. The general case is obtained by a composition of cases (3), (2), and (1). For symmetrical initial distributions there exist only two branches which join at $q = 1$, and the probability function may be written in a symmetrical form. When $n = 2$, the distribution of q for a symmetrical distribution is symmetrical about zero and invariant under a reciprocal transformation, and if the initial distribution possesses no moments and does not vanish at $x = 0$, the density of probability becomes infinite at $q = 0$. The distribution of q is not affected by changes in scale but is very sensitive to changes in origin. For a uniform distribution, the extremal quotient of a nonnegative variate has just the opposite qualities of the extremal quotient of a nonpositive variate. For variates changing sign, the extremal quotient is asymptotically negative.

31. The Distributions of the t and F Statistics for a Class of Nonnormal Populations. RALPH A. BRADLEY, Virginia Polytechnic Institute.

Series expansions of the cumulative distribution functions of t and of F in powers of t^{-1} and F^{-1} are obtained. The general method of derivation presented is valid for populations with density functions, $f(u)$, such that $f(u) > 0$, $f(u)$ is continuous, and has continuous derivatives for all values, $-\infty < u < \infty$. The coefficients of terms in these expansions are reduced from integrals, of multiplicity equal to the sample size, to products of coefficients, common to all populations of the class defined above, and integrals of no greater multiplicity than the number of groups of observations in the sample. Selected values of the common coefficients are given as well as illustrative examples for the Cauchy and "squared hyperbolic secant" population.

32. Note on the Behavior of the Characteristic Function of a Random Variable at Zero. M. ROSENBLATT, University of Chicago.

Let X be a random variable with characteristic function $\phi(z)$. Let $X_n = X$ when $|X| < n^{1/\alpha}$ and let $X_n = 0$ when $|X| \geq n^{1/\alpha}$. The following theorems are proved: (1) $1 - \phi(z) = o(|z|^\alpha)$, $0 < \alpha < 1$, at $z = 0$ if and only if $n \cdot Pr(|X| > n^{1/\alpha}) = o(1)$. (2) $1 - \phi(z) = o(|z|^\alpha)$, $1 \leq \alpha < 2$, at $z = 0$ if and only if $n \cdot Pr(|X| > n^{1/\alpha}) = o(1)$ and $E(X_n) = o(1)$. The results are obtained by making use of W. Feller's necessary and sufficient conditions for the weak law of large numbers (see W. FELLER, *Acta Univ. Szeged*, Vol. 8 (1937), pp. 191–201).

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest.

Personal Items

Dr. R. R. Bahadur, who received his Ph.D. in mathematical statistics from the University of North Carolina in June, 1950, is now an instructor in the Committee on Statistics of the University of Chicago.

Dr. T. A. Bancroft, Associate Professor of Statistics, Iowa State College, has been appointed Head of the Department of Statistics and Director of the Statistical Laboratory at Iowa State College.

Dr. Geoffrey Beall, formerly with the Research Laboratory of Swift & Co., Chicago, Illinois, has accepted a professorship in statistics in the Department of Mathematics, University of Connecticut, Storrs.

Dr. Archie Black has resigned his position as government statistician to devote more time to his work as Treasurer and Mathematical Consultant with the Mechanical Research Corporation of Chicago.

Professor David Blackwell of Howard University has been appointed Visiting Professor of Statistics at Stanford University for 1950–51.

Mr. Nils Blomqvist of the University of Stockholm has been appointed as Instructor in Mathematics and Statistics at Boston University for the academic year of 1950–51.

Professor Roque G. Carranza has started a course of twelve lectures on fundamentals of probability and statistics at Colegio Libre de Estudios Superiores in Buenos Aires which is a private nonprofit institution for higher studies.

Dr. Andrew L. Comrey, formerly an Assistant Professor of Psychology at the University of Illinois, has accepted a position as Assistant Professor of Psychology and Public Administration, Department of Psychology, University of Southern California, Los Angeles.

Dr. D. R. Cowan, who has been conducting many research projects for the coal industry and for various steel, food, paint, appliance, oil and other companies, has been appointed Professor of Marketing in the School of Business Administration, University of Michigan, Ann Arbor.

Dr. Robert E. Greenwood has been recalled to active duty by the Navy Department and is on leave from the Department of Applied Mathematics of the University of Texas.

Dr. Leo A. Goodman, who was a Social Science Research Council Research Training Fellow in the Mathematical Statistics Section of Princeton University, is now an Assistant Professor teaching statistics in the Sociology Department of the University of Chicago.

Mr. W. C. Hoffman is a thesis fellow at the Institute for Numerical Analysis during 1950–51.

Professor Paul Horst has returned to the Department of Psychology, University of Washington, Seattle. He took a year's leave of absence during 1949–1950 to be Director of Research in the Educational Testing Service at Princeton, New Jersey.

Dr. Stanley L. Isaacson, who was a Naval Research Assistant at Columbia University, has been appointed Assistant Professor of Statistics in the Statistical Laboratory at Iowa State College, effective September, 1950, where he will be employed in research and teaching. Dr. Isaacson spent last summer working as a Mathematical Statistician with the Operations Research Office in Washington, D. C.

Dr. Allyn W. Kimball has resigned his position as Experimental Statistician at the USAF School of Aviation Medicine, Randolph Air Force Base, and will be on the Mathematics Panel at the Oak Ridge National Laboratory.

Mr. William Kruskal is now associated with the Committee on Statistics at the University of Chicago.

Professor E. L. Lehmann is on leave of absence from the University of California, Berkeley, for the academic year 1950–51. He is teaching at Columbia University during the first semester and at Princeton University the second.

Mr. Garnet McCreary was awarded at the Commencement June 15, 1950, the degree of Doctor of Philosophy in Statistics at Iowa State College. His dissertation was entitled "Cost Functions for Sample Surveys." He has been appointed Assistant Professor in the Department of Mathematics, University of Manitoba, Winnipeg, Canada, effective September 1950, where he will be employed in teaching, consulting and research.

Professor William B. Michael, who held the position of Assistant Professor of Psychology at Princeton University for the past three years, accepted an appointment as Associate Professor of Psychology at San Jose State College commencing September, 1950.

Dr. Stanley W. Nash, formerly Associate in Mathematics at the University of California, Berkeley, has accepted an appointment as Assistant Professor of Mathematics and Research Consultant in Statistics at the University of British Columbia at Vancouver, effective July 1, 1950.

Dr. H. W. Norton has resigned as statistician of the Accountability Branch of the Atomic Energy Commission to become Professor of Agricultural Statistics in the Illinois Agricultural Experiment Station and the University of Illinois.

Dr. A. E. Paull, formerly biometrician for the Grain Research Laboratory, Board of Grain Commissioners, Winnipeg, Canada, has accepted a position as Associate Statistician in the Department of Statistical Research, Abitibi Power & Paper Company, Limited, Toronto, Canada.

Dr. Paul Peach, formerly Associate Professor at the Institute of Statistics, University of North Carolina, is now head of the Data Analysis Branch, Test Department at the Naval Ordnance Test Station, China Lake, California.

Dr. Raymond P. Peterson, formerly a research fellow at the Institute for Numerical Analysis, National Bureau of Standards, Los Angeles, received his doctorate from University of California at Los Angeles in June and has accepted a position as Instructor of Mathematics at the University of Washington, Seattle.

Dr. P. Ratoosh, formerly a lecturer at Columbia University, is now an Instructor in the Department of Psychology at the University of Wisconsin.

Mr. Joseph S. Rhodes has accepted a position as Mathematical Statistician in the office of the United States Air Force Comptroller, Washington, D. C. He is acting in the capacity of Mathematical Advisor to the Director of Statistical Services on the design of sample surveys.

Miss Rosemary Savey has accepted a position as Instructor in Statistics and Research Assistant in the Bureau of Business Research, University of Toledo.

Dr. Esther Seiden, formerly lecturer and research fellow at the University of California, Berkeley, accepted an assistant professorship in the Department of Statistics, School of Business Administration, University of Buffalo, New York.

Abraham Wald

Abraham Wald, a Fellow of the Institute, and Mrs. Wald were killed in a plane crash in India on December 13, 1950. Professor Wald was born in Cluj, Romania, October 31, 1902. His academic training was received at the University of Cluj (L.M., 1927) and the University of Vienna (Ph.D., 1931). He was a research associate of the Austrian Institute for Business Cycle Research until 1938 when he came to the United States. He had since been associated with Columbia University, becoming Professor of Mathematical Statistics and Executive Officer of the Department of Mathematical Statistics. Professor Wald made important contributions to the fields of mathematical statistics and probability, pure mathematics, and mathematical economics. He had written almost 100 papers in these fields as well as two books, *Sequential Analysis* and *Statistical Decision Functions*.

Professor Wald had been a president of the Institute, a member of the Council of the Institute and of the Editorial Board of the *Annals*, a vice-president and Fellow of the American Statistical Association, and a Fellow of the Econometric Society.

Statistics Summer Session at Virginia Polytechnic Institute

The Department of Statistics, Virginia Polytechnic Institute, will hold a special summer session August 8-25, 1951, for graduate students, research workers, and technicians in government and industry. Special emphasis will be given to statistics in economics and engineering. Several visiting professors will participate in the lecturing. For details write to the Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

The following Ph.D. degrees with major in mathematical statistics were granted at the University of North Carolina in 1950:

Name	Thesis	Minor
Raghu Raj Bahadur	On a Class of Decision Problems in the Theory of R Populations	Experimental Statistics and Mathematics
Kenneth A. Bush	Orthogonal Arrays	Mathematics and Economics
Max Halperin	Estimation in Truncated Sampling Processes	Experimental Statistics and Mathematics
Sharad-Chandra S. Shrikhande	Construction of Partially Balanced Designs and Related Problems	Experimental Statistics
Shanti A. Vora	Bounds on the Distribution of Chi-square	Experimental Statistics

New Members

The following persons have been elected to membership in the Institute.

(September 1, 1950 to November 30, 1950)

Berger, Richard, M.A. (Columbia Univ.), Statistician-Economist, General Aniline & Film Corporation, 230 Park Avenue, New York 17, New York.

Boll, C. H., B.S. (Stanford Univ.), Graduate Student in Statistics, Stanford University, *1247 Cowper, Palo Alto, California.*

Chacon, Enrique, S. J., Ph.D. (Univ. of Madrid), Professor of Statistics, University of Deusto, *Apartado 1, Bilboa, Spain.*

Curcio, F. L., M.S. (Univ. of Pa.), Graduate Student, Department of Mathematical Statistics, Columbia University, *559 Broadlawn Terrace, Vineland, New Jersey.*

Derman, Cyrus, A.M. (Univ. of Pa.), Graduate Student, Department of Mathematical Statistics, Columbia University, *449 MacDade Boulevard, Collingdale, Pennsylvania.*

de Finetti, Bruno, Ph.D. (Univ. of Milano), Professor, University of Trieste, *via Coroneo 43, Trieste, Italy.*

Geisser, Seymour. B.A. (City College of N. Y.), Graduate Student, University of North Carolina, *B-Dormitory, Room 312, Chapel Hill, North Carolina.*

Getchell, B. C., Ph.D. (Univ. of Mich.), Research Analyst, Department of Defense, *903 N. Wayne St., Apt. 305, Arlington, Virginia.*

Guerreiro, Amaro, B. Litt. (Oxford Univ.), Head of Research Division, Instituto Nacional de Estatistica, *48, Av. Marques de Tomar, Lisbon, Spain.*

Haddad, R. K., B.A. (N. Y. Univ.), Teaching Assistant in Psychology, Graduate School of Arts and Sciences, New York University, *43-02 63 Street, Woodside, New York.*

Hogg, R. V., Ph.D. (Univ. of Iowa), Assistant Professor, Department of Mathematics and Astronomy, State University of Iowa, Iowa City, Iowa.

Inselmann, E. H., A.B. (Temple Univ.), Graduate Student, Columbia University, *4253 N. 6th St., Philadelphia 40, Pennsylvania.*

Kitagawa, Tosio, Ph.D. (Tokyo Univ.), Professor of Mathematical Statistics, Kyushu University, and Chief, Committee of Research Association of Statistical Sciences, Faculty of Science, University of Kyushu, Fukuoka, Japan

Kurtz, T. E., A.B. (Knox College, Ill.), Research Assistant, Mathematics Department, Princeton University, Fine Hall, Box 708, Princeton, New Jersey.

Lantell, Gunnar, Fil. Kand. (Lund), Actuary of Försäkringsaktiebolaget Hansa Stockholm 7, Sweden.

Leimbacher, W. R., Dipl. Math. (Switzerland), Teaching Assistant at Federal Institute of Technology, *64, Ringstresse, Zurich 57, Switzerland.*

McCall, C. H., Jr., A.B. (George Wash. Univ.), Assistant in Statistics and Graduate Student, George Washington University, *6701-44th Street, Chevy Chase 15, Maryland.*

Matérn, Bertil, Fil. Lic. (Stockholm Univ.), Assistant Professor, Swedish Forest Research Institute, *Lappkarrsvagen 47, Stockholm 50, Sweden.*

Matthias, R. H., A.B. (Amherst College), Graduate Student, Department of Mathematical Statistics, University of North Carolina, *Purefoy Road, Chapel Hill, North Carolina.*

Ortiz, C. L. B., Ingeniero Civil (Paris), Asesor Analista, Direccion Navional de Estadistica, Contraloria General de la Republica, Bogota, Colombia.

Poch, F. A., Licencia do en Ciencias (Univ. of Madrid), Official of Instituto Nacional de Estadistica; Specialist of Section of Methodology; Assistant Professor of Mathematical Statistics, University of Madris, *Federico Rubio, 106, Madrid, Spain.*

Sastry, N.S.R., Ph.D. (London School of Econ.), Officer and Director of Statistics, Department of Research and Statistics, Reserve Bank of India, Post Bag No. 1036, Bombay 1, India.

Skibinsky, Morris, B.S. (City College of N. Y.), Graduate Student, Department of Mathematical Statistics, University of North Carolina, *Room 323 B-Dormitory, Chapel Hill, North Carolina.*

Sullivan, J. R., M.A. (Georgetown Univ.), Instructor in Mathematics, Clemson College, South Carolina (on leave); and Graduate Student, University of North Carolina, *18-C, Lennox, Chapel Hill, North Carolina.*

Tornqvist, Leo, Ph.D. (Abo Akademi), Professor in Statistics, Institute of Statistics, University of Helsinki, Helsinki, Suomi (Finland).

von Guerard, Hermann W., Director, Statist. Amts., Lambertusstr. 1, Dusseldorf, Germany.

Wunsche, Gunther, Dipl. Math. (Tech. Univ. of Dresden), Chefmathematiker und Prokurist, Universitatsdozent, *Lenbachplatz 4, Munich 2, Germany.*

REPORT OF THE CHICAGO MEETING OF THE INSTITUTE

The thirteenth Annual Meeting and forty-fifth meeting of the Institute of Mathematical Statistics was held in Chicago, December 27–29, 1950. Headquarters were at the Congress Hotel. Sessions were held at the Congress Hotel, Roosevelt College and the Palmer House. One or more sessions were held in conjunction with one or more of the following organizations: the American Statistical Association, the Econometric Society, the American Association of University Teachers of Insurance, the American Economic Association, the American Farm Economic Association, the American Marketing Association, the American Psychological Association, the American Public Health Association, the American Society for Quality Control (Chicago Section), the Association for Computing Machinery, the Biometric Society (Eastern North American Region), the Population Association of America, and the Psychometric Society. The following 266 members of the Institute attended:

Helen Abbey, F. S. Acton, Beatrice Aitchison, A. A. Alchian, J. E. Alman, R. L. Anderson, T. W. Anderson, E. E. Ard, K. J. Arnold, K. J. Arrow, Max Astrachan, G. J. Auner, R. R. Bahadur, E. W. Bailey, J. C. Bain, T. A. Bancroft, E. W. Barankin, Walter Bartky, W. D. Baten, R. E. Bechhofer, B. M. Bennett, Richard Berger, Z. W. Birnbaum, C. I. Bliss, Isadore Blumen, C. R. Blyth, A. H. Bowker, R. A. Bradley, Dorothy S. Brady, A. E. Brandt, M. F. Bresnahan, C. A. Bridger, Jean Bronfenbrenner, I. D. J. Bross, R. W. Burgess, L. D. Calvin, J. M. Cameron, E. S. Cansado, A. G. Carlton, C. G. Carlyle, O. S. Carpenter, Maria Castellani, F. R. Cella, Herman Chernoff, Randolph Church, W. G. Cochran, C. H. Coombs, Jerome Cornfield, J. H. Cover, Gertrude M. Cox, C. C. Craig, E. L. Crow, S. L. Crump, E. E. Cureton, Cuthbert Daniel, D. A. Darling, Besse B. Day, F. R. Del Priore, D. B. DeLury, W. E. Deming, B. W. Dempsey, Lucile Derrick, J. L. Doob, H. F. Dorn, D. B. Duncan, C. W. Dunnett, David Durand, A. M. Dutton, P. S. Dwyer, Churchill Eisenhart, Lila Elveback, Benjamin Epstein, H. P. Evans, W. D. Evans, W. T. Federer, Robert Ferber, J. W. Fertig, Evelyn Fix, M. M. Flood, E. J. Frank, L. R. Frankel, D. A. S. Fraser, H. A. Freeman, H. C. Fryer, R. P. Gage, M. A. Girshick, Mary A. Goins, L. A. Goodman, Roe Goodman, B. G. Greenberg, S. W. Greenhouse, J. A. Greenwood, L. E. Grosh, F. A. Gross, F. E. Grubbs, Harold Gulliksen, L. S. Gunlogson, John Gurland, R. K. Haddad, R. J. Hader, K. W. Halbert, Max Halperin, F. J. Halton, M. H. Hansen, H. H. Harman, T. E. Harris, H. L. Harter, Mina Haskind, P. M. Hauser, W. C. Healy, Jr., F. M. Hemphill, J. L. Hodges, Jr., William Hodgkinson, Jr., R. G. Hoffmann,

J. F. Hofmann, R. V. Hogg, H. B. Horton, D. G. Horvitz, Harold Hotelling, E. E. House-man, W. G. Howard, C. J. Hoyt, Leonid Hurwicz, P. E. Irick, S. L. Isaacson, J. E. Jackson, C. M. Jaeger, A. T. James, E. H. Jebe, R. J. Jessen, H. L. Jones, L. B. Kahn, Leo Katz, L. S. Kellogg, H. J. Kelly, Oscar Kempthorne, Nathan Keyfitz, A. W. Kimball, Jr., E. P. King, L. R. Klein, L. A. Knowler, Lila F. Knudsen, C. F. Kossack, R. L. Kozelka, K. H. Kramer, William Kruskal, Solomon Kullback, T. E. Kurtz, R. A. Leibler, H. O. Levine, G. J. Lieberman, Gilbert Lieberman, J. E. Lieberman, R. F. Link, Michel Loève, G. F. Lunger, W. G. Madow, C. J. Maloney, John Mandel, Nathan Mantel, E. S. Marks, Mary Marquardt, Jacob Marschak, Margaret P. Martin, Pat Maxwell, Jr., K. O. May, P. J. McCarthy, G. E. McCreary, D. C. McCune, P. W. McGann, F. E. McIntyre, Brock-way McMillan, Margaret Merrill, Robert Mirsky, A. M. Mood, R. H. Morris, J. E. Morton, L. E. Moses, Jack Moshman, Frederick Mosteller, B. D. Mudgett, Hugo Muench, M. R. Neifeld, C. J. Nesbitt, Jerzy Neyman, R. T. Nichols, M. L. Norden, J. I. Northam, H. W. Norton, G. B. Oakland, E. G. Olds, P. S. Olmstead, Bernard Ostle, Toby Oxtoby, A. E. Paull, M. P. Peisakoff, B. E. Phillips, Frank Proschan, Joan E. Raup, L. J. Reed, J. S. Rhodes, P. R. Rider, B. A. Rojas, C. F. Roos, S. N. Roy, M. M. Sandomire, F. E. Satter-thwaite, L. J. Savage, Henry Scheffé, M. A. Schneiderman, Elizabeth L. Scott, R. H. Shaw, R. W. Shephard, Jack Sherman, W. A. Shewhart, I. H. Siegel, Jack Silber, P. B. Simpson, Rosedith Sitgreaves, H. F. Smith, J. H. Smith, Milton Sobel, Herbert Solomon, L. D. Sommers, F. A. Sorensen, Mortimer Spiegelman, E. W. Stacy, B. R. Stauber, R. G. D. Steel, C. M. Stein, H. W. Steinhaus, F. F. Stephan, O. F. Stewart, J. V. Sturtevant, B. R. Suydam, Zenon Szatrowski, J. V. Talacko, Dan Teichroew, J. G. C. Templeton, B. J. Tepping, D. J. Thompson, G. R. Treanor, A. E. Treloar, J. W. Tukey, G. W. Tyler, S. A. Tyler, S. A. Vora, D. F. Votaw, Jr., Helen M. Walker, D. L. Wallace, W. A. Wallis, F. A. Weck, Samuel Weiss, M. E. Wescott, Eric Weyl, Phillips Whidden, S. S. Wilks, C. P. Winsor J. Wolfowitz, M. A. Woodbury, Holbrook Working, W. J. Youden, R. K. Zeigler.

At 10 a.m., Wednesday, December 27, 1950 the American Statistical Association joined the Institute in one of two sessions held at that time for contributed papers. Albert H. Bowker of Stanford University presided. The following papers were presented:

1. *Cost Functions for Sample Surveys. Preliminary Report.* Garnet E. McCreary, University of Manitoba and Iowa State College.
2. *On a Preliminary Test for Pooling Mean Squares in the Analysis of Variance.* A. E. Paull, Abitibi Power & Paper Company, Limited, Toronto, Canada.
3. *Estimation for Sub-Sampling Designs Employing the County as a Primary Sampling Unit.* Emil H. Jebe, Iowa State College and North Carolina State College.
4. *The Probability Distribution of the Number of Isolates in a Social Group.* Leo Katz, Michigan State College.
5. *Estimating Population Size Using Sequential Sampling Tagging Methods.* Leo A. Goodman, University of Chicago.
6. *Application of the Distribution of a Linear Form in Chi-square Variates.* Arthur Grad and Herbert Solomon, Office of Naval Research, Washington, D. C.
7. *A Large Sample t-statistic which is Insensitive to Nonrandomness.* (By Title.) John E. Walsh, The Rand Corporation.

At the second session for contributed papers held at 10 a.m., Wednesday, December 27, 1950, K. J. Arnold of the University of Wisconsin presided. The following papers were presented:

8. *Conditional Expectation and Convex Functions.* E. W. Barankin, University of California, Berkeley.

9. *Transformation Parameters*. Melvin P. Peisakoff, The Rand Corporation.
10. *A Generalization of the Neyman-Pearson Fundamental Lemma*. Henry Scheffé, Columbia University.
11. *Nonparametric Estimation V, Sequentially Determined Statistically Equivalent Blocks*. D. A. S. Fraser, University of Toronto.
12. *A Bayes Approach to a Quality Control Model*. M. A. Girshick and Herman Rubin, Stanford University.
13. *On the Translation Parameter Problem for Discrete Variables*. David Blackwell, Stanford University.
14. *On Ratios of Certain Algebraic Forms*. Robert V. Hogg, State University of Iowa.
15. *The Economics of Sampling*. (By Title.) Norman Rudy, Sacramento State College.
16. *Exact Tests of Serial Correlation Using Noncircular Statistics*. (By Title.) G. S. Watson, University of Cambridge, and J. Durbin, London School of Economics.
17. *Stochastic Difference Equations with a Continuous Time Parameter. Preliminary Report*. (By Title.) S. G. Ghurye, University of North Carolina.
18. *Nonsequential Problems in the Case of k Hypotheses. Preliminary Report*. (By Title.) Herman Chernoff, University of Illinois.

Also at 10 a.m., Wednesday, December 27, 1950, the Institute joined the American Statistical Association (Biometrics Section) and the Biometric Society (Eastern North American Region) in a session on *Statistical Problems in Radio-Biology*. A. E. Brandt of the United States Atomic Energy Commission was chairman. The papers presented were *Gene Mutations in Populations* by Bruce Wallace of the Long Island Biological Laboratory, *Long-Term Radiation Experiment in Dogs* by S. Lee Crump of the University of Rochester, and *Metabolism of Labeled Carbon Compounds* by Hardin B. Jones of the University of California at Berkeley. The papers were discussed by H. Fairfield Smith of the University of North Carolina and C. W. Sheppard of the Oak Ridge National Laboratory.

At 2 p.m., Wednesday, December 27, 1950, the American Statistical Association and the American Society for Quality Control (Chicago Section) joined the Institute in a session devoted to an address, *Statistical Control*, by W. A. Shewhart of the Bell Telephone Laboratories. E. G. Olds of the Carnegie Institute of Technology presided.

Also at 2 p.m., Wednesday, December 27, 1950, the Institute joined the American Statistical Association (Biometrics Section), the American Farm Economic Association, the Biometric Society (Eastern North American Region), and the Psychometric Society in a session on *Theory of Variance Components*. W. J. Youden of the National Bureau of Standards was chairman. The papers presented were *The Present Status of Variance Component Analysis* by S. Lee Crump of the University of Rochester, *Testing a Linear Relation Among Variances* by William G. Cochran of Johns Hopkins University, and *Application to Regression and to Errors of Measurement* by John W. Tukey of Princeton University. The papers were discussed by A. M. Mood of the Rand Corporation.

Also at 2 p.m., Wednesday, December 27, 1950, the Institute joined the American Statistical Association and the American Association of University Teachers of Insurance in a session on *Developments in Actuarial Science*. Cecil J. Nesbitt of the University of Michigan was chairman. The papers presented were *Survey of Theoretical Developments* by Charles A. Spoerl of the Aetna Life Insurance

Company, and *Survey of Practical Applications* by E. A. Lew and Frank Weck of the Metropolitan Life Insurance Company. The papers were discussed by Alfred Guertin of the American Life Convention, Chicago.

At 4 p.m., Wednesday, December 27, 1950, the American Statistical Association and the Econometric Society joined the Institute in a session devoted to a *Half-Century of Progress* address, *Multivariate Analysis*, by T. W. Anderson of Columbia University. M. A. Girshick of Stanford University presided.

Also at 4 p.m., Wednesday, December 27, 1950, the Institute joined the American Statistical Association (Biometrics Section), American Society for Quality Control (Chicago Section), and the Biometric Society (Eastern North American Region) in a session on *Precision of Measurements*. W. Edwards Deming of the Division of Statistical Standards was chairman. The papers presented were *The Specification of Precision of Measurements* by Churchill Eisenhart of the National Bureau of Standards, *The Estimation of Precision of Measurements* by Frank E. Grubbs of the Aberdeen Proving Grounds, and *Estimate of Precision of Textile Instruments* by John C. Whitwell of Princeton University. The papers were discussed by H. Fairfield Smith of the University of North Carolina.

Also at 4 p.m., Wednesday, December 27, 1950, the Institute joined the American Statistical Association (Business and Economic Statistics Section), the American Economic Association, and the Econometric Society in a session on *Analysis of Choices Involving Risk*. Jacob Marschak of the Cowles Commission for Research in Economics was chairman. The papers presented were *Alternative Approaches to Theory of Choice in Risk-Taking Situations* by Kenneth J. Arrow of Stanford University and *An Experimental Measurement of Utility* by Frederick Mosteller of Harvard University. The papers were discussed by Armen Alchian of the University of California at Los Angeles and Franco Modigliani of the University of Illinois.

At 10 a.m., Thursday, December 28, 1950, the American Statistical Association joined the Institute in a session devoted to a *Half-Century of Progress* address, *Non-Parametric Inference*, by A. M. Mood of the Rand Corporation. P. S. Dwyer of the University of Michigan presided.

Also at 10 a.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association and the American Society for Quality Control (Chicago Section) in the first session on *Engineering*. Frederick J. Halton, Jr., of Deere & Company, was chairman. The paper presented was *Statistics in Production and Inspection* by Edwin G. Olds of the Carnegie Institute of Technology. The paper was discussed by Warren E. Jones of Desplaines, Illinois, and Charles A. Bicking of Hercules Powder Company.

Also at 10 a.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association (Section on the Training of Statisticians), the American Psychological Association, and the Psychometric Society in a session on *Statistical Literacy in the Social Sciences*. The address was given by Helen M. Walker of Columbia University. Philip M. Hauser of the University of Chicago was chairman.

Also at 10 a.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association (Biometrics Section) and the Biometric Society (Eastern North American Region) in a session on *Statistical Methods in Pharmacology and Immunology*. Lloyd C. Miller, director of Revision of the United States Pharmacopeia, was chairman. The papers presented were *Collaborative Bioassays* by Lila F. Knudsen, Food and Drug Administration, and *Statistical Methods in Immunology* by Herbert C. Batson of the Army Medical Center. The papers were discussed by Everett Welker of the American Medical Association and George Hunt of Bristol Laboratories.

At 2 p.m., Thursday, December 28, 1950, the American Statistical Association and the Econometric Society joined the Institute in a session devoted to an address, *Some Recent Advances in the Theory of Decision Functions*, by Jacob Wolfowitz of Columbia University. J. L. Doob of the University of Illinois presided.

Also, at 2 p.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association and the Population Association of America in a session on *Developments in United States Census Taking*. W. F. Ogburn of the University of Chicago was chairman. The papers presented were *Role of Research in Census Taking* by Morris Hansen, Bureau of the Census, *Evaluation of Census Results* by Eli Marks, Bureau of the Census, and *Census Programs and Operations* by A. Ross Eckler, Bureau of the Census. The papers were discussed by Nathan Keyfitz of the Dominion Bureau of Statistics, Ottawa, and Vergil D. Reed of the J. Walter Thompson Company.

Also at 2 p.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association, the American Psychological Association, and the Psychometric Society in a session on *Statistical Problems and Psychological Theory*. Allen Edwards of the University of Washington was chairman. The papers presented were *Statistical Problems and Psychological Scaling* by Clyde H. Coombs, University of Michigan, and *Statistical Problems and Learning Theory* by Kenneth W. Spence, State University of Iowa. The papers were discussed by Harold P. Bechtoldt, Iowa City, and Harold Gulliksen of the Educational Testing Service.

Also at 2 p.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association (Biometrics Section) and the Biometric Society (Eastern North American Region) in a session on *Applications of Variance Components*. G. W. Snedecor of Iowa State College was chairman. The papers presented were *Variance Components as a Tool for the Analysis of Sample Data* by Walter A. Hendricks, United States Department of Agriculture, *Consistency of Estimates of Variance Components* by R. E. Comstock and H. F. Robinson of North Carolina State College, and *Use of Components of Variance in Preparing Schedules for the Sampling of Baled Wool* by J. M. Cameron of the National Bureau of Standards. The papers were discussed by Walter T. Federer of Cornell University.

Also at 2 p.m., Thursday, December 28, 1950, the Institute joined the Ameri-

can Statistical Association and the American Society for Quality Control (Chicago Section) in the second session on *Engineering*. W. Edwards Deming of the Division of Statistical Standards was chairman. The papers presented were *Statistics in Engineering Research and Development* by Ellis R. Ott of Rutgers University, and *Statistical Developments in South Africa* by H. S. Sichel of the Educational Testing Service.

At 4 p.m., Thursday, December 28, 1950, the Psychometric Society and the Econometric Society joined the American Statistical Association (Section on the Training of Statisticians) and the Institute in a session devoted to a *Half-Century of Progress* address and a *Special Invited Paper, Statistical Inference*, by Jerzy Neyman of the University of California at Berkeley. S. N. Roy of the University of North Carolina presided.

Also at 4 p.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association (Biometrics Section, Business and Economic Statistics Section), the American Farm Economic Association, and the Biometric Society (Eastern North American Region) in a session on *Sample Survey Techniques*. W. F. Callander of Gainesville, Florida, was chairman. The papers presented were *Double Sampling and the Curtis Impact Study* by D. S. Robson of Cornell University and Arnold J. King of National Analysts, Inc., Philadelphia, *Approaches to Agricultural Price Statistics* by F. E. McVay and Henry Tucker of North Carolina State College, and *Problems in Rural Surveys* by R. L. Anderson and A. L. Finkner of North Carolina State College. The papers were discussed by B. R. Stauber of the Division of Agricultural Price Statistics and B. J. Tepping of the Bureau of the Census.

Also at 4 p.m., Thursday, December 28, 1950, the Institute joined the American Statistical Association, the Association for Computing Machinery, and the Psychometric Society in a session devoted to a *Round Table: What Can High Speed Electronic Computing Equipment Do For and To Statistics?* William G. Madow of the University of Illinois was moderator. Electronic Engineer Sam N. Alexander of the National Bureau of Standards and Expert User Byron Schreiner of the A. C. Nielson Company, Chicago, were the speakers. The papers were discussed by Howard C. Grieves of the Bureau of the Census and John J. Finelli of the Metropolitan Life Insurance Company.

At 10 a.m., Friday, December 29, 1950, the American Statistical Association joined the Institute in a session devoted to a *Half-Century of Progress* address, *Surveys*, by W. G. Madow of the University of Illinois. F. F. Stephan of Princeton University presided.

Also at 10 a.m., Friday, December 29, 1950, the Institute joined the Econometric Society in a session on *Problems of Incorrect and Incomplete Specification*. Merrill M. Flood of the Rand Corporation was chairman. The papers presented were *Some Specification Problems and Their Applications to Econometric Models* by Leonid Hurwicz of the University of Illinois and the Cowles Commission for Research in Economics, and *An Approach to Effects of Non-Normality*

in Tests of Significance by William Kruskal of the University of Chicago. The papers were discussed by T. W. Anderson of Columbia University and John W. Tukey of Princeton University.

Also at 10 a.m., Friday, December 29, 1950, the Institute joined the American Statistical Association, the Psychometric Society, and the American Psychological Association in a session on *Factor Analysis as a Statistical Tool*. The address was given by L. L. Thurstone of the University of Chicago. The paper was discussed by E. E. Cureton of the University of Illinois. Harold Gulliksen of the Educational Testing Service was chairman.

Also at 10 a.m., Friday, December 29, 1950, the Institute joined the American Statistical Association and the American Society for Quality Control (Chicago Section) in a session on *Statistics in the Physical Sciences*. The address was given by Walter Bartky of the University of Chicago. The paper was discussed by J. L. Doob of the University of Illinois. S. S. Wilks of Princeton University was chairman.

Also at 10 a.m., Friday, December 29, 1950, the Institute joined the American Statistical Association (Biometrics Section), the Biometric Society (Eastern North American Region), and the American Public Health Association in a session on *Statistical Methods in Medicine*. Hugo Muench of Harvard University was chairman. The papers presented were *A Stochastic Model of Relapse, Death and other Risks Following a Treatment* by Evelyn Fix and J. Neyman of the University of California, Berkeley, *The Design of Physiological and Clinical Investigations* by Donald Mainland of New York University and J. W. Hopkins of the National Research Council, Ottawa, and *Discriminatory Analysis* by Joseph L. Hodges, Jr., of the University of California, Berkeley. The papers were discussed by Samuel W. Greenhouse of the National Cancer Institute.

At 2 p.m., Friday, December 29, 1950, the American Statistical Association joined the Institute in a session devoted to an address, *Elements of Information Theory*, by Brockway McMillan of the Bell Telephone Laboratories. Solomon Kullback of George Washington University presided.

Also at 2 p.m., Friday, December 29, 1950, the Institute joined the Econometric Society and the American Economic Association in a session on *Collection and Use of Survey Data*. J. Neyman of the University of California at Berkeley was chairman. The papers presented were *Sample Surveys of Households: A New Tool in Econometrics* by Lawrence R. Klein of the University of Michigan, and *Use of Sample Surveys of Business Expectations and Plans* by Franco Modigliani of the University of Illinois. The papers were discussed by Paul F. Lazarsfeld of Columbia University.

Two sessions for contributed papers were held at 4 p.m., Friday, December 29, 1950. At one of these Oscar Kempthorne of Iowa State College presided. The following papers were presented:

19. *The Moments of a Multinormal Distribution after One-sided Truncation of Some or All Coordinates*. Z. W. Birnbaum and Paul L. Meyer, University of Washington.

20. *An Algorithm for the Determination of all Solutions of a Two-Person Zero Sum Game with a Finite Number of Strategies.* H. Raiffa, G. L. Thompson, and R. M. Thrall, University of Michigan.
21. *A Note on the Convolution of Uniform Distributions.* Edwin G. Olds, Carnegie Institute of Technology.
22. *On the Consistency of Certain Estimates of the Linear Structural Relation.* Elizabeth L. Scott, University of California, Berkeley.
23. *A 3-Decision Problem Concerning the Mean of a Normal Population.* R. R. Bahadur, University of Chicago.
24. *Consistent Estimate of the Slope of a Linear Structural Relation.* J. Neyman, University of California, Berkeley, and Charles M. Stein, University of Chicago.
25. *A Remark on Almost Sure Convergence.* Michel Loève, University of California, Berkeley.
26. *A Significance Test for Differences Among Ranked Treatments in an Analysis of Variance.* D. B. Duncan, Virginia Polytechnic Institute.

At the fourth session for contributed papers, the second at 4 p.m., Friday, December 29, 1950, W. D. Baten of Michigan State College presided. The following papers were presented:

27. *On Information and Sufficiency.* S. Kullback, George Washington University, and R. A. Leibler, Washington, D. C.
28. *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes.* T. W. Anderson, Columbia University, and D. A. Darling, University of Michigan.
29. *The Effect of Preliminary Tests of Significance on the Size and Power of Certain Tests of Univariate Linear Hypotheses with Special Reference to the Analysis of Variance. Preliminary Report.* Robert E. Bechhofer, Columbia University.
30. *The Exact Distribution of the Extremal Quotient.* E. J. Gumbel, New York, and L. H. Herbach, Columbia University. (The paper was read by J. A. Greenwood.)
31. *The Distributions of the t and F Statistics for a Class of Nonnormal Populations.* Ralph A. Bradley, Virginia Polytechnic Institute.
32. *Note on the Behavior of the Characteristic Function of a Random Variable at Zero.* M. Rosenblatt, University of Chicago. (Introduced by L. J. Savage.)

Also at 4 p.m., Friday, December 29, 1950, the Institute joined the American Statistical Association (Section on the Training of Statisticians) in an *R. A. Fisher Survey* session. Gertrude Cox of the University of North Carolina was chairman. The papers presented were *Revolution of Methods in Experimentation* by W. J. Youden of the National Bureau of Standards, and *The Impact of R. A. Fisher on Statistics* by Harold Hotelling of the University of North Carolina.

Also at 4 p.m., Friday, December 29, 1950, the Institute joined the American Statistical Association (Business and Economic Statistics Section), the American Marketing Association, the American Psychological Association, and the Psychometric Society in a session on *Measurement of Opinion*. William G. Cochran of Johns Hopkins University was chairman. The papers presented were *Implications for Factor Analysis of Lazarsfeld's Latent Structure Theory* by Bert J. Green of Princeton University, discussed by Paul F. Lazarsfeld of Columbia University, *A New Approach to Thurstone's Method of Scaling* by Frederick Mos-

teller of Harvard University, discussed by L. L. Thurstone of the University of Chicago, and *A Critical Analysis of Guttman's Theory of Principal Components in Attitude Measurement* by Philip J. McCarthy of Cornell University.

A meeting of the 1950 Council was held on Wednesday, December 27, 1950, at 12:00 noon, Professor J. L. Doob presiding. The Annual Business Meeting was held on Wednesday, December 27, 1950, at 7:00 p.m., Professor J. L. Doob presiding. A meeting of the 1951 Council was held on Friday, December 29, 1950, at 12:00 noon, Professor P. S. Dwyer presiding. The report of the Annual Business Meeting appears elsewhere in this issue.

<div align="right">

K. J. Arnold
Associate Secretary

</div>

MINUTES OF THE ANNUAL MEMBERSHIP MEETING, CHICAGO, DECEMBER 27, 1950

The meeting was called to order at 7:10 p.m. by President J. L. Doob. The annual reports of the President, Editor, and Secretary-Treasurer were read. They are printed elsewhere in this issue.

The Acting Secretary moved that Article 2 of the By-Laws of the Institute be amended so that the first two sentences read: "Members shall pay ten dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter Members shall pay ten dollars annual dues, of which seven dollars shall be for a subscription to the Official Journal." and that exception D be amended to read: "Any Member who resides outside the United States and Canada shall pay seven dollars annual dues." The motion carried.

The President asked for instructions from the membership as to the procedure to be followed in filling the unexpired term of Abraham Wald. It was voted that the candidate for the Council receiving the fifth largest number of votes be declared elected for a term of one year.

J. W. Tukey moved that it is the sense of this meeting that a four day annual meeting is preferable to a three day meeting even if this means meeting alone on the fourth day. The motion carried.

Harold Hotelling moved the adoption of the following resolution:

Whereas the death of Professor Abraham Wald, who with Mrs. Wald was killed in an airplane crash in India, deprives statistics of a vigorous, brilliant, and original contributor to its fundamental ideas; and

Whereas the future of statistical methods will be vitally affected by Abraham Wald's introduction of a formalized and accurate method of sequential analysis, and by his work on the foundations of statistical inference, including particularly the consideration of loss and risk functions, of general decision problems, of the minimax principle and the related theory of games, of the nature of the estimation of unknown quantities, and of the testing of hypotheses; and

Whereas the efforts of American industry and the military and naval services of supply were materially aided in the successful conduct of the Second World War by widespread application of Abraham Wald's work, particularly to the quality control of manufactured articles; and

Whereas his contributions to statistical methods and theory were substantial in such varied fields as the foundations of probability, inequalities on distributions in terms of moments, the treatment of time series, long cycles resulting from repeated integration, tolerance limits, analysis of variance, asymptotic large-sample distributions, and the estimation of parameters of stochastic processes; and

Whereas Abraham Wald contributed also to economics and economic statistics by his penetrating studies of equations of production and of general equilibrium, of index numbers of cost of living, and of the determination of indifference loci by means of Engel curves; and

Whereas he made in his earlier career in Europe valuable contributions to pure mathematics in the fields of differential geometry and the axiomatization of metric spaces; and

Whereas he served the American Statistical Association as Vice President and the Institute of Mathematical Statistics as President and as member of its Council and of the Editorial Board of the *Annals of Mathematical Statistics*; and

Whereas great inspiration is to be derived from the example of Abraham Wald in his surmounting of the difficulties caused by the discrimination and restrictions that, in his East European environment, denied him the opportunities of the primary and secondary schools; in his entrance to the university in his native city of Klausenburg by examinations for which he had prepared himself; in his graduation with distinction and his brilliant graduate work at the University of Vienna; in his migration to this country at the time of the fall of Austria; in his fortitude in enduring the loss of his nearest relatives by the Nazi policy of genocide; in his devotion to our science and in his habits of hard work which brought a great volume of substantial contributions; and in his ability to be friendly and kind under the severest strains; now therefore

Be it resolved that the American Statistical Association and the Institute of Mathematical Statistics jointly record their deepest sorrow and regret at the untimely passing away in middle life of a great contributor, and at the further tragedy that his wife also was taken; and that this Association extends its sincere sympathy and good wishes to the bereaved relatives and particularly to the two young children who remain.

The resolution was adopted unanimously by a rising vote.

The meeting adjourned at 8:05 p.m.

After the meeting the tellers posted the results of the election as follows:

President-Elect	M. A. Girshick
Members of the Council for 1951–1953	Harald Cramér
	A. M. Mood
	Jerzy Neyman
	S. S. Wilks
Member of the Council for 1951	E. L. Lehmann

K. J. Arnold
Acting Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1950

The general affairs of the Institute marked time this year, as far as the President's office was concerned, but fortunately the Institute's fine condition is not critically dependent on presidential activity.

The situation at the University of California has caused considerable discussion and evoked strongly differing opinions from members of the Institute. While there was general agreement that a deplorable situation had been created by arbitrary actions of the regents, there was no agreement as to what, if anything, the Institute should do about it. (Several other scientific organizations, including the American Mathematical Society, have passed resolutions on the subject.) A committee, consisting of Paul S. Dwyer (chairman), M. A. Girshick, and Henry Scheffé, was appointed to report to the Council on the facts, and to make recommendations as to possible action.

Institute activities at the year's end were overshadowed by the news of the tragic deaths of Professor Wald and Mrs. Wald in India. It is needless to record here the central place of Wald's work in the progress of statistics in recent years. A proper appreciation will be published in a later issue of the Annals.

Five new Fellows were elected at the Christmas meeting: R. C. Bose, J. L. Hodges, Jr., O. Reiersøl, H. Rubin, L. J. Savage.

The following Nominating Committee was appointed to serve for the year 1951:

G. W. Brown, *Chairman*	G. E. Nicholson
K. J. Arrow	H. W. Norton
E. L. Lehmann	J. W. Tukey

The following is a list of committees of the Institute in 1950:

1. *Committee to Encourage Membership outside the United States*
   T. W. Anderson, *Chairman*   M. Loève
   C. C. Hurd   J. Marschak
2. *Committee on Tabulation*
   H. O. Hartley, *Chairman*   C. C. Hurd
   C. I. Bliss   A. N. Lowan
   F. W. Dresch   W. G. Madow
   C. Eisenhart   H. G. Romig
   H. H. Germond   L. E. Simon
3. *Committee on the Directory*
   J. W. Tukey, *Chairman*
   C. Eisenhart
4. *Committee on Statisticians in the Government Service*
   W. E. Deming, *Chairman*
   C. Eisenhart
5. *Committee for the Christmas Meeting*
   W. G. Madow, *Chairman*   J. E. Morton
   M. H. Hansen   J. Wolfowitz
   M. Loève

6. *Committee for the 1950 Spring Meeting in North Carolina*
   H. Hotelling, *Chairman*   S. B. Littauer
   D. Blackwell   D. F. Votaw, Jr.
   H. Geiringer   S. S. Wilks
7. *Midwest Program Committee*
   O. Kempthorne, *Chairman*   K. May
   L. Hurwicz   L. J. Savage
   W. G. Madow   D. R. Whitney
8. *West Coast Program Committee*
   A. M. Mood, *Chairman*   W. J. Dixon
   Z. W. Birnbaum   J. L. Hodges, Jr.
   A. H. Bowker   P. G. Hoel
9. *Program Committee for the Oak Ridge Meeting*
   O. Kempthorne, *Chairman*   H. W. Norton
   A. S. Householder   L. J. Savage
   H. Levene
10. *Committee to Look into Less Expensive Possibilities for Printing the Annals*
    T. W. Anderson, *Chairman*   P. S. Dwyer
    W. E. Deming   S. S. Wilks
11. *Committee for Special Invited Papers*
    W. G. Madow, *Chairman*   O. Kempthorne
    T. W. Anderson   A. M. Mood
12. *Committee to Revive the Statistical Research Memoirs*
    H. Scheffé, *Chairman*   C. C. Hurd
    T. W. Anderson   G. Kuznets
    W. Bartky
13. *Representative of the I. M. S. to the American Association for the Advancement of Science*
    J. Neyman
14. *Representative of the I. M. S. to the National Research Council, Division of Physical Sciences*
    W. Bartky
15. *Representative of the I. M. S. to the Mathematical Policy Committee*
    H. Scheffé
16. *Representative of the I. M. S. to the Joint Committee for Development of Statistical Applications in Engineering and Manufacturing*
    B. Epstein
17. *Representatives of the I. M. S. to the Inter-Society Committee on the Mathematical Training of Social Scientists*
    W. G. Madow, *Chairman*
    T. W. Anderson
18. *Representative of the I. M. S. to the American Academy of Political and Social Science*
    F. F. Stephan

                                             J. L. Doob
                                             President

December 27, 1950

REPORT OF THE SECRETARY-TREASURER OF THE
INSTITUTE FOR 1950

At the beginning of 1950 the Institute had 1164 members and during the period covered by this report 129 new members (6 of whom were appointed to membership by Institutional Members) joined the Institute and 19 were reinstated. During 1950 the Institute lost 73 members of which 26 were by resignation, 46 were cancelled for non-payment of dues, and one member was deceased. Judging from the information available at this date, the Institute will have 1239 members as it starts 1951.

During the year, a list of statisticians in countries outside of the United States, compiled by the Committee to Encourage Membership outside the United States headed by T. W. Anderson, was solicited for membership by this office. Thirty-three of the new members were obtained primarily because of this solicitation, and it is possible that some few additional membership applications may still result from this campaign.

Meetings of the Institute held during 1950 included those at Chapel Hill, North Carolina on March 17–18; at Chicago, Illinois on April 28–29; at Berkeley, California on August 5; and at Chicago on December 27–29, 1950. The Secretary wishes to call attention to the excellent work of Professor W. G. Madow, Program Chairman for the December, 1950 annual meeting and of the members who served as Assistant and Associate Secretaries at these meetings: Professor Herbert E. Robbins at Chapel Hill; Professor K. J. Arnold at Chicago in April and December; and Professor L. J. Savage at Chicago.

The following Fellows served as members of the Committee on Fellows: Henry Scheffé, Chairman, T. W. Anderson, M. A. Girshick, E. L. Lehmann, H. E. Robbins, and F. F. Stephan.

INSTITUTE OF MATHEMATICAL STATISTICS
Statement of Condition
December 31, 1950

ASSETS

Bank		$3,569.24
Dues Receivable		95.00
Subscriptions Receivable		897.20
U.S. Government Bonds		4,888.00
Total Assets		$9,449.44

LIABILITIES AND RESERVES

Amount Due to Printer	$2,145.00	
Withholding Tax Payable	106.20	
Miscellaneous Liabilities	75.25	$2,326.45
Reserve for Dues Advanced	345.00	
Reserve for Subscriptions Advanced	2,159.66	
Reserve for Life Members	2,757.50	5,262.16
Total Liabilities and Reserves		$7,588.51
*Surplus (Excess of Assets over Liabilities)		1,860.83
		$9,449.44

* Surplus is not adjusted for inventory of back issues estimated at a nominal value of $17,621.00 (67¢ per issue).

Neither is the surplus adjusted for the reserve for reprinting back issues. There are four issues of which we have less than 25 copies on hand that will have to be printed early in 1951 at an estimated cost of $1,300, and there are four other issues of which we have between 26 and 50 copies which will probably have to be reprinted later in the year.

<div align="center">Revenue and Expense Statement</div>

<div align="center">For the year ending December 31, 1950</div>

Revenues

Dues Revenue	$9,087.70	
Subscriptions Revenue	3,875.48	
Sale of Back Issues	5,465.81	
Interest Earned on Bonds	100.00	
Miscellaneous Revenue	124.94	$18,653.93

Expenses

Printing of *Annals* Current	$8,774.81	
Reprinting of Back Issues	1,313.20	
Salary Expense	3,000.00	
Miscellaneous Printing of Stationery and Postage	891.77	
Contributions to American Mathematical Society	244.13	
Miscellaneous Office Expense	229.55	
Editorial Expense	200.00	
Meeting Expense	77.45	
Binding Expense	35.00	$14,765.91
Excess of Revenues over Expenses		$ 3,888.02
Excess of Liabilities over Assets December 31, 1949		2,027.19
Excess of Assets over Liabilities December 31, 1950		$ 1,860.83

It has been our practice to set up an amount equal to all life membership payments as a liability and to hold all these funds in reserve until the death of the member—after which his payment is released to the general fund. There were no new life membership payments during 1950 nor were there any deaths among life members. The total number of members therefore remains as 32.

In the last annual report, the membership was advised that the accounting system of the Institute was going to be completely re-organized and placed on a modern basis. The form of the preceding report results from this change. We no longer take credit for prepaid dues and subscriptions for the coming year in our report of revenue and expenses for the year just closed. We have followed the advice of our accounting consultants in not including among our assets the very non-liquid inventory of back issues of the *Annals*.

During 1950, we had a remarkable sale of back issues amounting to $5465.81, an increase of approximately $2100 over the previous year and $2500 over 1948. Furthermore, we reprinted only four issues during the year, whereas we reprinted twelve issues in each of the two preceding years. This extremely favorable conjunction of items more than compensated for the excess of our normal

operating cost over our normal operating income from dues and subscriptions. It is very unlikely that such a situation will prevail again, at least in the immediate future. The present international situation will probably cut our back number sales appreciably. Furthermore, we now have twelve issues in relatively short supply which may have to be reprinted in 1951. We were fortunate in 1950, but it would be expecting too much to count on a windfall from back numbers to take up the slack resulting from inadequate dues and subscription rates.

<div align="right">CARL H. FISCHER
<i>Secretary-Treasurer</i></div>

December 20, 1950

REPORT OF THE EDITOR OF THE ANNALS FOR 1950

The past year has seen a new editorial organization for the *Annals* instituted. The Constitution of the Institute of Mathematical Statistics adopted in 1948 provides for the election of Associate Editors by the Council. In the new organization, the Associate Editors not only collaborate with the Editor on matters of policy but also assume a large share of the responsibility for consideration of manuscripts submitted to the *Annals*. In establishing policy and procedure, the Editorial Committee has relied heavily on the study made by the Committee on Editorial Policy of the *Annals*.

The Editorial Committee wishes to acknowledge the cooperation of the previous Editor, S. S. Wilks, in the inauguration of the new editorship. The front cover of the *Annals* now bears recognition of Professor Wilks' accomplishments during his twelve years as Editor of the *Annals*, in accordance with the resolution passed by the Institute of Mathematical Statistics at its 1949 membership meeting.

The 1950 volume of the *Annals* contained 56 papers of which 22 were notes. The number of pages printed, 624, was about the same as in the several preceding years. The need for an increased number of pages for the *Annals* continues. Unfortunately, during the past year the cost of printing has risen approximately 10%. It is unavoidable that the budget for the *Annals* be increased.

"Fundamental Limit Theorems of Probability Theory" by M. Loève, published in the September, 1950, issue, was the first paper invited by the Special Invited Papers Committee. It is expected that by the invitation of this committee more expository and review papers will be provided for the *Annals*.

On behalf of the Editorial Committee, the Editor takes this opportunity to acknowledge the generous refereeing assistance of the following: F. C. Andrews, F. J. Anscombe, Kenneth Arrow, E. W. Barankin, Robert Bechhofer, Agnes Berger, Z. W. Birnbaum, C. Blyth, Albert Bowker, D. G. Chapman, H. Chernoff, Randal H. Cole, Allen T. Craig, C. C. Craig, D. A. Darling, W. J. Dixon, H. F. Dodge, S. G. Ghurye, H. R. J. Grosch, Frank E. Grubbs, Leon Herbach, J. L. Hodges, Jr., W. Hoeffding, E. L. Kaplan, J. L. Kelley, William Kruskal,

P. J. McCarthy, Paul Meier, R. J. Monroe, R. B. Murphy, G. E. Noether, Edward Paulson, Melvin Peisakoff, John Riordan, H. G. Romig, Herman Rubin, L. J. Savage, W. L. Scott, E. Seiden, R. G. D. Steel, Charles Stein, D. F. Votaw, Jr., John E. Walsh, Ransom Whitney.

The Editor is indebted to David Bruce Hanchett and Jack Laderman for preparation of manuscripts for the printer and to Miss Jean Hanson for other editorial and office assistance.

<div align="right">

T. W. ANDERSON
Editor

</div>

December 8, 1950

PUBLICATIONS RECEIVED

The Editor has received a number of books from publishers for review purposes. Because of the present shortage of space in the *Annals* due to the large number of papers submitted and because several other journals carry on a broad reviewing service (particularly *Mathematical Reviews* and the *Journal of the American Statistical Association*), the Editorial Committee has decided that at this time *Annals* space cannot be devoted to reviews of these books. For the information of the readers, the publications received will be listed.

Acceptance Sampling (A Symposium), American Statistical Association, Washington, D. C., 1950, iv + 155 pp., $1.50.

Anuario Estadistico de España, (Instituto Nacional de Estadistica) Presidencia del Gobierno, Madrid, 1950, lv + 898 pp.

FELLER, WILLIAM, *An Introduction to Probability Theory and Its Applications*, Vol. 1, John Wiley and Sons, Inc., New York, 1950, xii + 419 pp., $6.00.

GEBELEIN, H., *Zahl und Wirklichkeit, Grundzüge einer Mathematischen Statistik*, 2nd ed., Quelle and Meyer, Heidelberg, 1949, xii + 430 pp.

Measurement and Prediction, (Studies in Social Psychology in World War II, Vol. 4), Princeton University Press, Princeton, 1950, x + 756 pp., $10.00.

MOOD, ALEXANDER McFARLANE, *Introduction to the Theory of Statistics*, McGraw-Hill Book Company, Inc., New York, 1950, xiii + 433 pp., $5.00.

NEYMAN, J., *First Course in Probability and Statistics*, Henry Holt and Company, New York, 1950, ix + 350 pp., $3.50.

Table of Bessel Functions $Y_0(z)$ and $Y_1(z)$ for Complex Arguments, (Computation Laboratory, National Bureau of Standards) Columbia University Press, New York, 1950, xl + 427 pp., $7.50.

Contents

ASA

The American Statistical Association invites as members all persons interested in:
1. development of new theory and method
2. improvement of basic statistical data
3. application of statistical methods to practical problems.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS FOR 1951

President:

P. S. Dwyer, University of Michigan, Ann Arbor

President-Elect:

M. A. Girshick, Stanford University, Stanford

Secretary-Treasurer:

C. H. Fischer, University of Michigan, Ann Arbor

Editor:

T. W. Anderson, Columbia University, New York

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote cooperation between the field of pure research and the fields of application.

Membership dues including subscription to the ANNALS OF MATHEMATICAL STATISTICS are $10.00 per year within the United States and Canada and $7.00 per year elsewhere. Dues and inquiries regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.

MEETINGS OF THE INSTITUTE

SPRING MEETING—OAK RIDGE, TENNESSEE—March 15-17, 1951.
To be held in conjunction with the Biometric Society.

SANTA MONICA, CALIFORNIA—June 18, 1951.
Abstracts must be in the hands of Associate Secretary J. L. Hodges, Jr., Statistical Laboratory, University of California, Berkeley 4, California, not later than May 1.

SUMMER MEETING—MINNEAPOLIS, MINNESOTA—September 7-8, 1951.
To be held in conjunction with the American Mathematical Society and the Mathematical Association of America. Abstracts must be in the hands of Associate Secretary K. J. Arnold, North Hall, University of Wisconsin, Madison, Wisconsin, not later than July 15.